Pedosphere **29**(4): 421–431, 2019
doi:10.1016/S1002-0160(19)60811-8
ISSN 1002-0160/CN 32-1315/P

Published by Elsevier B.V. and Science Press

PEDOSPHERE

www.elsevier.com/locate/pedosphere

Removing Moisture Effect on Soil Reflectance Properties: A Case Study of Clay Content Prediction

Yaron OGEN[1,3,*], Shira FAIGENBAUM-GOLOVIN[2], Amihai GRANOT[3], Yoel SHKOLNISKY[2], Naftaly GOLDSHLEGER[4] and Eyal BEN-DOR[3]

[1] *The Porter School of Environmental Studies, Tel Aviv University, Tel Aviv 6997801 (Israel)*
[2] *Department of Applied Mathematics, Tel Aviv University, Tel Aviv 6997801 (Israel)*
[3] *Remote Sensing Laboratory, Department of Geography and the Human Environment, Tel Aviv University, Tel Aviv 6997801 (Israel)*
[4] *Department of Civil Engineering, Ariel University, Ariel 407000 (Israel)*

(Received July 9, 2018; revised August 29, 2018)

ABSTRACT

Visible, near-infrared and shortwave-infrared (VNIR-SWIR) spectroscopy is an efficient approach for predicting soil properties because it reduces the time and cost of analyses. However, its advantages are hampered by the presence of soil moisture, which masks the major spectral absorptions of the soil and distorts the overall spectral shape. Hence, developing a procedure that skips the drying process for soil properties assessment directly from wet soil samples could save invaluable time. The goal of this study was twofold: proposing two approaches, partial least squares (PLS) and nearest neighbor spectral correction (NNSC), for dry spectral prediction and utilizing those spectra to demonstrate the ability to predict soil clay content. For these purposes, we measured 830 samples taken from eight common soil types in Israel that were sampled at 66 different locations. The dry spectrum accuracy was measured using the spectral angle mapper (SAM) and the average sum of deviations squared (ASDS), which resulted in low prediction errors of less than 8% and 14%, respectively. Later, our hypothesis was tested using the predicted dry soil spectra to predict the clay content, which resulted in R^2 of 0.69 and 0.58 in the PLS and NNSC methods, respectively. Finally, our results were compared to those obtained by external parameter orthogonalization (EPO) and direct standardization (DS). This study demonstrates the ability to evaluate the dry spectral fingerprint of a wet soil sample, which can be utilized in various pedological aspects such as soil monitoring, soil classification, and soil properties assessment.

Key Words: dry spectral fingerprint, nearest neighbor spectral correction, partial least squares, reflectance spectra, soil moisture, soil property, spectroscopy, wet soil

Citation: Ogen Y, Faigenbaum-Golovin S, Granot A, Shkolnisky Y, Goldshleger N, Ben-Dor E. 2019. Removing moisture effect on soil reflectance properties: A case study of clay content prediction. *Pedosphere*. **29**(4): 421–431.

INTRODUCTION

Reflectance spectra across the visible, near-infrared and shortwave-infrared (VNIR-SWIR) region (400–2 500 nm) can potentially offer a reliable tool for quantitative and qualitative assessment of soil properties (Chang *et al.*, 2001; Summers *et al.*, 2011). In the last two decades, much attention has been devoted in many scientific fields to extracting information from reflectance measurements (Chabrillat *et al.*, 2002, 2013; Dunn *et al.*, 2002; Cohen *et al.*, 2005; Christy, 2008; Zornoza *et al.*, 2008; Volkan Bilgili *et al.*, 2010; Viscarra Rossel and Webster, 2012; Arslan *et al.*, 2014), mostly in the laboratory domain. In addition, various statistics-based methods have been developed to model the relationship between the spectral reflectance and soil constituents (Cozzolino and Morón, 2006; Janik *et al.*, 2009; Reeves III and Smith, 2009). Ben-Dor and Banin (1995), Islam *et al.* (2003), Viscarra Rossel *et al.* (2006), and Wijewardane *et al.* (2016) examined the prediction of clay content from a dry spectrum and obtained coefficient of determination (R^2) and root mean square error of prediction (RMSEP) of 0.56 and 10.3, 0.05 and 9.8, 0.41 and 2.3, and 0.44 and 12.54, respectively.

Although reflectance spectroscopy can provide reliable predictions of soil properties, it is fairly sensitive to the influence of moisture, as demonstrated in many studies (*e.g.*, Muller and Décamps, 2001; Liu *et al.*, 2002; Wu *et al.*, 2009; Zhu *et al.*, 2010; Minasny *et al.*, 2011; Nocita *et al.*, 2013; Ge *et al.*, 2014; Rienzi *et al.*, 2014; Castaldi *et al.*, 2015; Ji *et al.*, 2015; Jiang *et al.*, 2016; Roberts and Cozzolino, 2016; Wijewardane *et al.*, 2016). Soil moisture affects the spectrum by de-

*Corresponding author. E-mail: yaronogen@gmail.com.

creasing the overall albedo across it and by amplifying water absorption at 1 400, 1 900 nm, and at the shoulder (2 300–2 500 nm) (Twomey *et al.*, 1986; Lobell and Asner, 2002; Ge *et al.*, 2014). As a result, soil moisture may change the soil's spectral fingerprint and blur significant absorption features caused by major soil chromophores such as organic matter (400–900 and 1 700 nm), clay minerals (1 400, 1 900, and 2 200 nm), and carbonate minerals (2 300 nm) (Clark, 1999).

In general, all models introduced to date for evaluating soil properties perform well with dry soil spectra under laboratory conditions. Accordingly, prior to applying those models, the samples must be brought to the laboratory and dried at ambient (room) temperature and humidity for several days. Obtaining a procedure that would enable measurement of soil properties using data taken under field conditions could save invaluable time (sample collection and drying) and allow experts to obtain the dry soil spectra instantly for *in-situ* assessment of soil properties.

This is extremely important when soil spectral libraries (SSLs) measured under dry and controlled conditions are projected onto hyperspectral remote-sensing images of fields with unknown moisture content. This was demonstrated in the recent paper by Castaldi *et al.* (2018), who applied the Land Use and Cover Area Frame Survey (LUCAS) SSL directly on the Airborne Prism Experiment (APEX) hyperspectral images to determine soil organic matter content without taking into account the moisture effect. Accordingly, the present study attempted to bridge the gap between airborne measurements that characterize field conditions (including moisture) and the SSL archive (room-dried), and also to demonstrate this big step for other SSLs, such as the global SSL of Viscarra Rossel *et al.* (2016). This emphasizes the need to deal with the soil moisture effect at the field level, in order to exploit the huge amount of data available in SSLs.

In recent years, several algorithms have been suggested to perform this task. The first widely used method is the external parameter orthogonalization-partial least squares (EPO-PLS) algorithm (Roger *et al.*, 2003; Wijewardane *et al.*, 2016). This method predicts soil properties from spectra that are sensitive to moisture effects, obtained under field (wet) conditions. Wijewardane *et al.* (2016) compared four statistical methods (PLS, random forest (RF), artificial neural network (ANN), and support vector machine (SVM)) in order to predict clay content and concluded that although EPO was applied on the data, no significant improvement was observed (R^2 = 0.02 and RMSEP = 14.45, using ANN). One of the main drawbacks of the EPO-PLS is its limited ability to remove the moisture effect from mineral-rich soils; accordingly, further studies are needed to investigate the extent to which mineral composition determines the algorithm's effectiveness (Ackerson *et al.*, 2015). Another method was suggested by Ji *et al.* (2015), who examined the effect of moisture on spectral information and proposed a direct standardization (DS) method to predict dry soil spectra from wet spectra. The DS algorithm finds the linear relation between the wet and dry soil spectra in a least-squares sense. A recent paper (Roudier *et al.*, 2017) compared the performance of the two methods, and concluded that although it was similar, the EPO outperformed the DS method when using a smaller number of samples.

Each of these methods has its weaknesses; *e.g.*, the EPO-PLS requires creating a model for each soil property, and DS is highly sensitive to noise. In this study, we used two independent approaches, PLS regression (Wold *et al.*, 2001) and nearest neighbor spectral correction (NNSC), to predict dry soil spectra by modeling the relationship between the wet and dry spectra for spectral correction. To construct the prediction models for dry soil spectra, a dataset containing spectra of dry soils along with their different water contents was obtained and utilized for training and testing purposes. The quality of the predicted dry spectra can influence the assessment of the soil property in question, so it was evaluated by the average sum of deviations squared (ASDS) (Ben-Dor *et al.*, 2004) and the spectral angle mapper (SAM) parameter (Kruse *et al.*, 1993). Clay content is one of the most important factors for water-retention capacity (Banin and Amiel, 1970). The ability to predict soil properties was then demonstrated by estimating clay content from the predicted (dry) soil spectra. We built a new model for this purpose using PLS_Toolbox 8.2.1 (2016) (Eigenvector Research, Inc., USA). The model was calibrated with additional soil samples and then applied to our predicted dry soil spectra. Prediction accuracy for clay content was assessed using R^2, RMSEP, the bias, the standard error of performance (SEP), the ratio of performance to deviation (RPD), and the slope.

MATERIALS AND METHODS

Soil spectral database

Soils of eight common types in Israel were sampled from 66 different locations, yielding a total of 162 dry soil samples with differing clay contents (Table I). All samples were air-dried, ground, and sieved (< 2 mm) before laboratory measurements.

Figure 1 shows the soil characteristics in the form of a ternary diagram, which provides information on the minimum and maximum percentages of clay (and the equivalent sand and silt percentages) for the different soil types. Each sample was scanned using an Analytical Spectral Device (ASD) laboratory reflectance spectrometer (ASD, USA) with a spectral range of 350–2 500 nm. All samples were measured using a contact probe equipped with a halogen bulb and Spectralon® (Labsphere, USA) as a white reference. For data analysis, we only used the spectral range of 400–2 450 nm as it shows a high signal-to-noise ratio.

To prepare the soils with varying moisture contents, each soil sample was divided into three subsamples that were subjected to one of the following processes, creating three new subgroups of samples: i) samples that have undergone a drying process (220 sub-samples), ii) samples that have undergone a wetting process (448 subsamples), and iii) samples without processing, *i.e.*, soils were measured in their initial dry state (162 sub-samples). Subgroups i and ii contained a total of 668 samples and subgroup iii contained 162 samples, resulting in a total of 830 soil samples. Table I summarizes the soil types, the number of spectra measured (for each soil sample while being wetted or dried), and the minimum, maximum, mean, and standard deviation (SD) of water contents.

Thereafter, the 668 samples were randomly partitioned into two datasets (A and B) with proportions of 85% and 15%, respectively. Dataset A contained 568 samples and was used for model construction, and dataset B contained 100 samples and was used for accuracy evaluation. Each sample from datasets A and B was associated with only one sample from the 162 ini-

TABLE I

Soil types based on Israeli classification and World Reference Base for Soil Resources (WRB) and water content data of the samples for spectra measuring using the wetting and drying processes

Israeli classification	WRB classification	No. of dry samples	No. of spectra (wetting)	No. of spectra (drying)	Total number of measured spectra	Water content[a)]			
						Min	Max	Mean	SD
						g kg^{-1}			
Brown-red sandy soil	Rhodic Luvisol	34	138	29	201	0	231	79.9	68.4
Alluvial soil	Eutric Fluvisol	18	47	4	69	0	527	107.0	104.0
Terra rossa	Epileptic Luvisol	19	50	63	132	0	600	92.0	120.0
Grumusol	Chromic Vertisol	52	132	31	215	0	442	123.0	81.2
Basaltic soil	Haplic Luvisol	15	28	8	51	0	233	122.0	61.6
Sierozem	Luvic Calcisol	20	35	5	60	0	148	68.2	43.0
Loessial sandy soil	Luvic Calcisol	2	12	42	59	0	455	148.0	122.0
Rendzina	Calcaric Leptosol	2	6	38	46	0	556	171.0	163.0
Total number		162	448	220	830				

[a)]Min = minimum; max = maximum; SD = standard deviation.

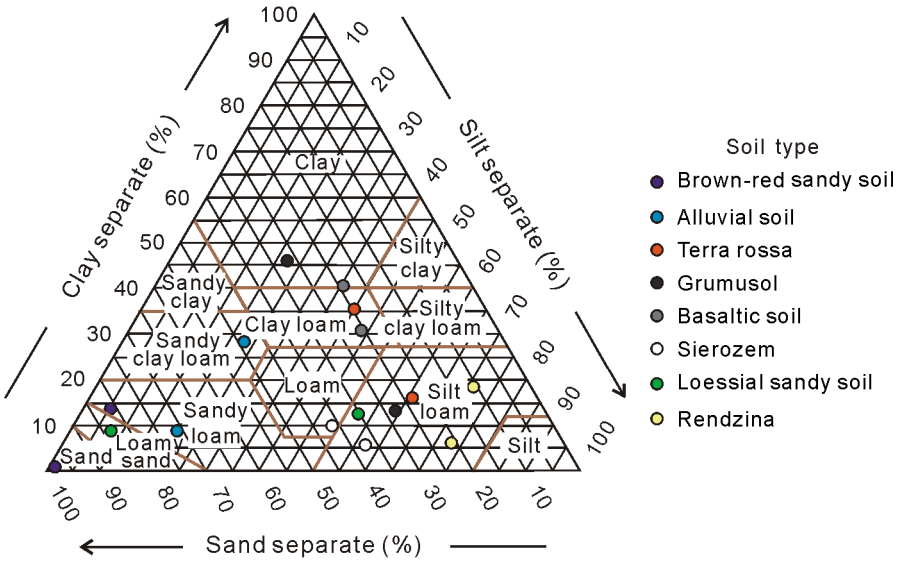

Fig. 1 Ternary diagram of two samples from each soil type (based on Israeli classification) with the minimum and maximum percentages of clay (and the equivalent sand and silt percentages).

tial soil samples (subgroup iii). This was done to ensure that all samples were paired and that every single sample from either the wetting or drying processes had an equivalent dry sample.

The samples had a height of approximately 2 cm and a weight of 100 g. For the drying process, each soil sample was saturated with distilled water and then air-dried under ambient conditions until only hygroscopic water remained (approximately 10 d). For the wetting process, a small amount of distilled water was added to the samples (using sprinklers), which were then stirred, weighed, and scanned. We continued to sprinkle water until the saturation point was reached. During the wetting process, all samples were weighed to evaluate their moisture content and a spectral measurement was taken daily.

Soil clay content database

In addition, dataset C was created using new soil samples for clay content prediction model construction and validation. We utilized this dataset to evaluate the accuracy of the predicted clay content from the predicted dry soil spectra. The dataset consisted of 213 air-dried soil samples, which were collected from diverse climatic and topographic regions in Israel. The percentage of clay was measured with a laser diffraction particle-size analyzer (Mastersizer 2000, Malvern Panalytical, UK). The statistical information of the datasets A, B, and C is provided in Table II.

TABLE II

Number of samples and the minimum, maximum, mean, and standard deviation values of clay percentages in datasets A, B, and C used for soil clay content prediction model construction, accuracy evaluation, and model validation, respectively

Dataset	No. of samples	Clay			
		Minimum	Maximum	Mean	Standard deviation
		%			
A	568	0.0	47.0	14.1	7.7
B	100	0.8	46.5	15.8	8.5
C	213	0.6	44.0	12.4	9.1

Using PLS_Toolbox, the PLS method was used to construct a clay content prediction model for dataset C. Figure 2 presents the clay-content model using 162 calibration samples and 51 validation samples, which resulted in $R^2 = 0.71$ and RMSEP = 4.93.

Mathematical procedures: prediction of dry soil spectra

Given a spectrum of a soil with unknown water content, S^{wet}, we would like to predict the spectrum of the soil under dry conditions, S^{dry}. This problem can be addressed by constructing a model that describes the relationship between the wet and dry soil spectra. Applying the model to wet soil spectra should result in spectra resembling the dry soil spectra. The accuracy of the constructed model can then be tested by comparing the predicted dry spectra with spectra measured directly on the dry soil samples. In this study, we proposed two methods to remove the effect of moisture from a given spectrum.

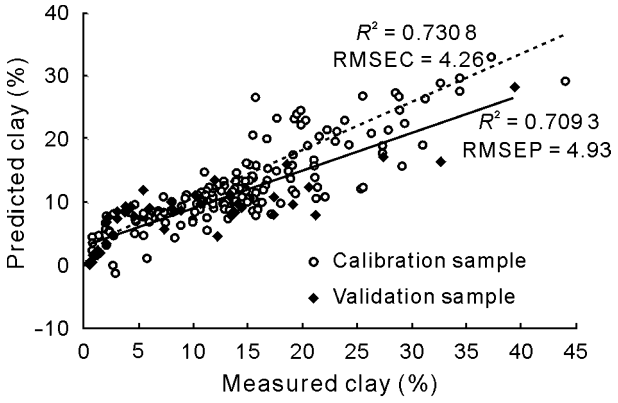

Fig. 2 Soil clay content prediction model constructed using the partial least squares (PLS) method for dataset C with 162 calibration and 51 validation samples. The dashed and solid lines are the regression lines for the calibration and validation samples, respectively. R^2 is coefficient of determination, RMSEC is root mean square error of calibration, and RMSEP is root mean square error of prediction.

Dry soil spectra prediction using PLS. To model the relationship between the dry and wet soil spectra, the commonly employed PLS algorithm (Wold *et al.*, 2001) was used. The PLS algorithm models the relations between sets of observations by means of latent variables, which are not always directly observed. Thus, PLS regression finds the observation components that best predict the latent variable. The underlying assumption is that the sought-after relation is linear. In our scenario, the construction of the PLS model used spectral data from the training dataset (dataset A). The wet soil spectra served as predictor variables, while the dry soil spectra were set as the response matrix. We constructed the PLS-2 model *via* MATLAB's plsregress function, with cross-validation, to prevent over-fitting (and to find the number of components). The created model is presented as a matrix of weights, $M_{i,j} \in \mathbb{R}$. Thus, the ith wavelength of the predicted dry spectrum is the weighted sum of the measured wavelengths of the wet soil spectrum, where the weights are the rows of M. Given a wet spectrum $S^{\text{wet}} \in \mathbb{R}^N$ (N is the number of wavelengths or 2 050 in our case), we can predict the ith wavelength of the dry soil spectrum ($\hat{S}_i^{\text{dry}}$) *via* Eq. 1:

$$\hat{S}_i^{\text{dry}} = S_1^{\text{wet}} + \sum_{j=2}^{N} M_{i,j} S_j^{\text{wet}} \quad (1)$$

where S_1^{wet} is the linear term in the prediction, $\sum_{j=2}^{N} M_{i,j}$ is the learned weight of the wet soil spectrum to predict the ith wavelength of the dry soil spectrum, and S_j^{wet} is the jth wavelength of the wet soil spectrum.

The main advantage of the PLS model is its flexibility. First, a single model can be used to predict dry spectra of various soil types (all of the types used to construct the model). Namely, the weight matrix M incorporates all of the information from all of the soils in the training set. Second, the model does not use any information related to the amount of water in the soil. Thus, once a soil is sampled in the field, we can instantly predict its dry spectrum. In addition, the values of M indicate the wavelengths of the wet soil spectrum that are most important for the process of predicting the dry soil spectrum. It should be noted that the PLS method uses all of the wavelengths, as opposed to the piecewise direct standardization method (PDS), in which only a few wavelengths are used for the prediction.

Dry soil spectra prediction using NNSC. The second method is the NNSC method, which transforms wet soil spectra into dry soil spectra by multiplying the former by a predefined coefficient vector. To calculate this coefficient vector, called the spectral correction coefficient, dataset A, which contains samples of wet soil spectra and their corresponding dry soil spectra, was used. The spectral correction coefficient vector, $\bar{c} \in \mathbb{R}^N$, is defined as the proportion between a wet soil spectrum and the corresponding dry soil spectrum from the training set (wavelength-by-wavelength division), as shown in Eq. 2:

$$\bar{c} = \frac{S^{\text{dry}}}{S^{\text{wet}}} \quad (2)$$

Subsequently, to locate the relevant spectral correction coefficient vector for a new wet soil spectrum (after normalization), we find the nearest neighbor spectrum from the reference spectral dataset A. The nearest neighbor method uses a spectral proximity, d, which integrates several spectrum-similarity measures. Specifically, the distance metric is the average value of four measurements: SAM and ASDS operating on the normalized spectra (using Eq. 1), ASDS using the original spectra, and the spectral gradient m, given by Eq. 3:

$$m = \frac{\text{Ref}_{800} - \text{Ref}_{500}}{800 - 500} \times 1\,000 \quad (3)$$

where Ref is the reflectance value at a given wavelength and the denominator is the wavelength in nanometers. Upon receiving a new wet soil spectrum $S_{\text{new}}^{\text{wet}}$, the spectrum is normalized as follows:

$$\bar{S} = S/\|S\| \quad (4)$$

where $\bar{S}$ is the normalized spectrum and S is the original spectrum. Hence, $\|\bar{S}\|^2 = 1$. Then, the nearest wet spectrum is located in the reference data using the spectral proximity measure:

$$k^* = \text{argmin}_k \text{d}(S_{\text{new}}^{\text{wet}}, \bar{S}_k^{\text{wet}}) \quad (5)$$

where k^* is the index of the closest wet spectrum from dataset A to the examined spectrum $S_{\text{new}}^{\text{wet}}$, k is a running index on all the elements in dataset A, $\bar{S}_k^{\text{wet}}$ is the spectrum of the wet soil with the index k from dataset A, and $\bar{S}_k^{\text{wet}} \in$ A. Later, the corresponding spectral correction coefficient vector $\bar{c}_{k^*}$ is located for the nearest neighbor spectrum and used to correct the wet spectrum (wavelength-by-wavelength multiplication) by using Eq. 6:

$$S_{\text{NNSC}}^{\text{dry}} = S_{\text{new}}^{\text{wet}} \times \bar{c}_{k^*} \quad (6)$$

where $S_{\text{NNSC}}^{\text{dry}}$ is the spectrum of the dry soil predicted utilizing the NNSC method.

Spectral fit assessment

Measuring the distance between two spectra is a fundamental step in evaluating the accuracy of our approaches. The spectral comparison (before and after normalization) is illustrated in Fig. 3. The normalized spectra can be compared by the SAM error ($\text{SAM}_{\text{error}}$, Eq. 7) and the ASDS (Eq. 8):

$$\text{SAM}_{\text{error}} = \cos(\text{Rr}, \text{Rm}) \quad (7)$$

$$\text{ASDS} = \frac{1}{N}\Big(\sum_{n=1}^{N} (\text{Rr}_n/\text{Rm}_n - 1)\Big) \quad (8)$$

where Rr and Rm are the reference spectrum and the measured spectrum, respectively, Rr_n is the value obtained from a given wavelength in the reference spectrum, and Rm_n is the value obtained from the same wavelength in the measured spectrum. Both SAM and ASDS represent accuracy measures; SAM represents the similarity between vectors, and ASDS is equivalent to the relative error.

Clay content prediction methods

In this study, we compared our methods with the EPO (Roger *et al.*, 2003) and DS (Ji *et al.*, 2015) me-

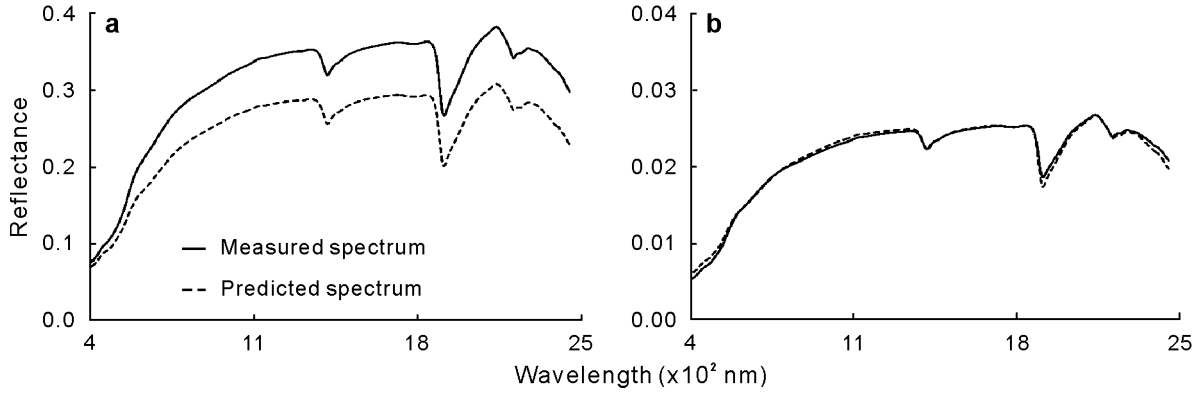

Fig. 3 Comparison between dry soil spectra (measured and predicted) before normalization (a) and those after normalization (b).

thods. Herein, we provide a short description of those methods and the error analysis after applying them on our datasets.

The EPO method is comprised of two training phases. First, a transformation, which performs dimensional reduction of the wet soil signature, is found *via* principal component analysis (PCA). Then, the relationship between the lower dimensional data and the clay content measurements is found *via* PLS. To support the two-phase training, two different datasets are needed. One is utilized for training the PCA model (containing information on wet and dry soil spectra (A1)), and the other is used for training the PLS model (containing wet soil spectra and the corresponding clay content data (A2)). In our case, dataset A (with 568 samples) was divided into two subsets, A1 and A2, each containing half of the samples (*i.e.*, datasets of 284 samples each). We implemented the EPO method in MATLAB and optimized its parameters following Wijewardane *et al.* (2016) to obtain the best possible prediction. Thus, we performed an exhaustive search to locate the best c in the EPO algorithm, which is the number of EPO dimensions. For each c, cross-validation of the PLS was implemented, and the optimal k was selected. We then chose the c that minimized the RMSEP. The numerical range of the tested parameter values was 1–30 (for both parameters). Thus, the parameters were chosen based on the training of 11 PLS factors and 17 EPO dimensions. We then constructed the EPO model, and tested it on dataset B.

The DS method finds the linear relation between the wet and dry soil spectra, in a least-squares sense (after centering the data). The DS method was implemented in MATLAB, and the linear relation between the wet and dry soil spectra was estimated using dataset A. For clay content prediction, we used the model constructed on dataset C. We estimated accuracy and precision of the models using the slope of the linear fit line between the measured and predicted values of clay content, where a slope value closer to 1 indicates a more ideal model and by using R^2, RMSEP, bias, SEP, and RPD (Esbensen *et al.*, 2002), which are given in Eqs. 9–12:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(Y_{\mathrm{predicted}} - Y_{\mathrm{measured}})^2}{n}} \tag{9}$$

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n}(Y_{\mathrm{predicted}} - Y_{\mathrm{measured}})}{n} \tag{10}$$

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(Y_{\mathrm{predicted}} - Y_{\mathrm{measured}} - \mathrm{bias})^2}{n-1}} \tag{11}$$

$$\mathrm{RPD} = \frac{\mathrm{SD}_{\mathrm{measured}}}{\mathrm{SEP}} \tag{12}$$

where $Y_{\mathrm{predicted}}$ is the predicted clay value, Y_{measured} is the measured clay value, n is the number of samples, and $\mathrm{SD}_{\mathrm{measured}}$ is the SD of the measured clay value.

RESULTS

Predicting the dry soil spectra

Upon PLS and NNSC model construction (using dataset A), we predicted the dry spectra of samples from dataset B and evaluated their accuracy. Figure 4 shows the predicted spectra for each soil type. Each graph contains the measured dry spectrum, the wet spectrum, and the two predicted dry soil spectra using the PLS and NNSC methods, respectively.

Table III summarizes the SAM, ASDS, and RMSE values used to compare the predicted dry soil spectra with the observed reference spectra for all soil types from dataset B. Both of the methods resulted in low prediction errors; *i.e.*, the mean of SAM and ASDS was less than 8% and 14%, respectively.

The results raised the question of whether water content has an impact on the accuracy of the prediction. As can be seen in Fig. 5, the SAM values were independent of the water content, and therefore our pre-

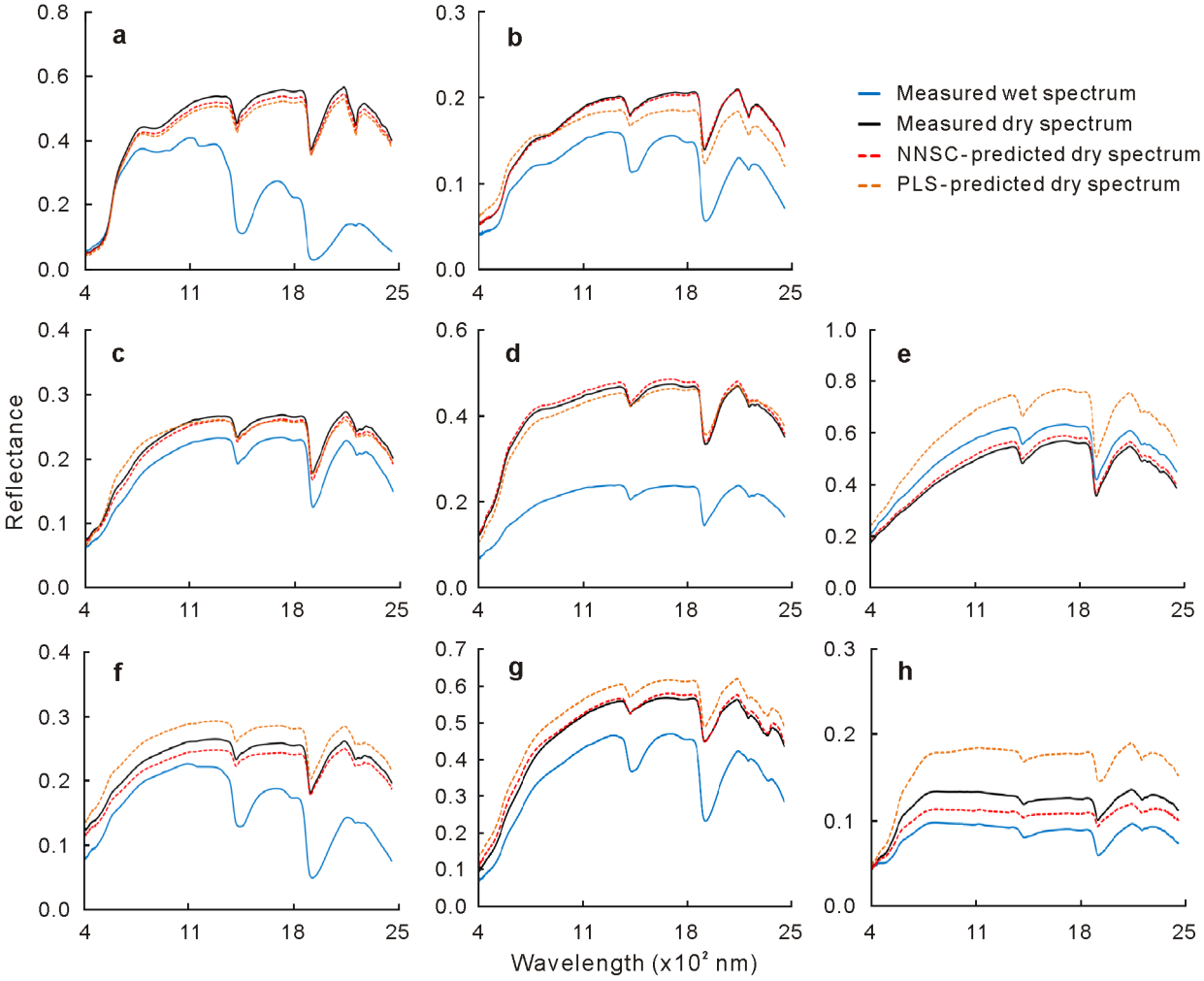

Fig. 4 Comparison of measured dry and wet soil spectra and the nearest neighbor spectral correction (NNSC)- and partial least squares (PLS)-predicted dry soil spectra for samples representative of the eight soil types in Israel based on Israeli classification: brown-red sandy soil (a), terra rossa (b), Grumusol (c), loessial sandy soil (d), rendzina (e), alluvial soil (f), sierozem (g), and basaltic soil (h).

TABLE III

Mean values of spectral angle mapper (SAM), average sum of deviations squared (ASDS), and root mean square error (RMSE) used to evaluate the quality of the predicted dry soil spectra using partial least squares (PLS), nearest neighbor spectral correction (NNSC), and direct standardization (DS)

Soil type based on Israeli classification	PLS			NNSC			DS		
	SAM	ASDS	RMSE	SAM	ASDS	RMSE	SAM	ASDS	RMSE
Brown-red sandy soil	0.038	0.009	0.000 018	0.095	0.285	0.085	0.050	0.046	0.118
Alluvial soil	0.056	0.008	0.000 027	0.061	0.097	0.036	0.063	0.016	0.139
Terra rossa	0.059	0.022	0.000 993	0.071	0.090	0.028	0.080	0.089	0.094
Grumusol	0.051	0.006	0.033 900	0.060	0.121	0.047	0.070	0.019	0.119
Basaltic soil	0.041	0.008	0.000 020	0.078	0.285	0.069	0.093	0.052	0.166
Sierozem	0.030	0.002	0.000 014	0.146	0.025	0.123	0.038	0.003	0.095
Loessial sandy soil	0.118	0.104	0.000 057	0.032	0.028	0.048	0.130	0.140	0.061
Rendzina	0.020	0.001	0.000 009	0.057	0.063	0.073	0.043	0.005	0.030
Total mean value	0.050	0.016	0.000 024	0.076	0.133	0.060	0.069	0.044	0.108

diction methods were not affected by specific water contents.

In addition, we analyzed the coefficient matrix for the PLS, EPO, and DS methods (except for the NNSC method which is data dependent). For that purpose, we picked the wavelengths whose coefficient sum was above the 99 percentile threshold. The most important spectral wavelengths for each method were as follows:

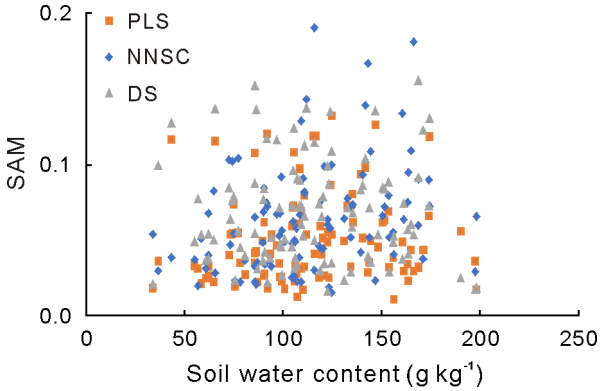

Fig. 5 Influence of soil water content on the spectral angle mapper (SAM) accuracy of the three dry soil spectra prediction methods of partial least squares (PLS), nearest neighbor spectral correction (NNSC), and the direct standardization (DS).

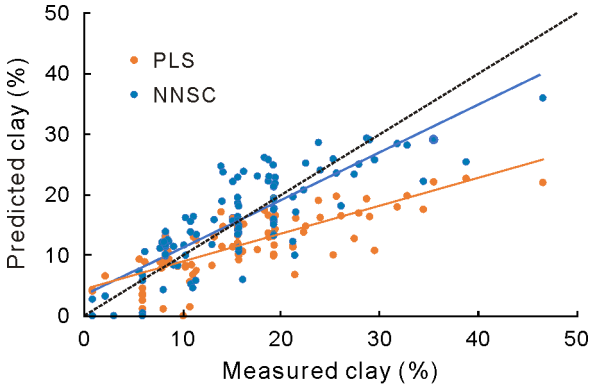

Fig. 6 Comparison of soil clay measured by laser diffraction and predicted by the partial least squares (PLS) and nearest neighbor spectral correction (NNSC) methods ($n = 100$). The dashed line represents the 1:1 line, and the yellow and orange solid lines are the regression lines for NNSC and PLS methods, respectively.

400–460, 1 340–1 350, 1 880, 1 920 nm for the PLS; 500–550, 630, 660, 680, 730, 854, 900, 950–1 000, 1 050, 1 080, 1 100, 1 930, 2 075 nm for the DS; and 1 990–2 020 nm for the EPO. These wavelengths are strongly related to the water absorption bands, as was studied previously by Ben-Dor and Banin (1995) and Pope and Fry (1997).

We compared our dry soil spectral prediction approaches with the DS method (Table III). As can be seen from Table III, the error rates of DS and PLS are close. However, the main weakness of DS is its sensitivity to noise. In order to demonstrate its instability, a short experiment was performed. We added Gaussian noise to the existing data and checked the accuracy of the two models. In statistics, a common assumption is that the data are sampled with Gaussian noise. Therefore, we added a 5-dB signal-to-noise ratio per sample into the dataset. The dry soil spectrum prediction results showed that the SAM error arising from the PLS increases from 1.6% (in clean conditions) to 3% (with noise), while the SAM error of the DS prediction increases from 4.4% to 37%. These results demonstrate that the DS is indeed much more sensitive to noise that may subsequently affect the clay content prediction.

Predicting soil reflectance properties: clay content as a case study

We next applied the predicted dry soil spectra obtained from the PLS and NNSC methods to predicting the clay content of each sample, as shown in Fig. 6. For comparison, we also applied the DS method on the same model and compared the clay content prediction accuracy with that received directly from the EPO method.

The best result was achieved using the PLS method, which attained $R^2 = 0.69$, SEP = 5.11, and RPD = 1.6. The EPO method achieved the best scores in RMSEP and bias, 5.23 and −0.06, respectively, followed by the NNSC, with slope = 0.72. The performance of DS did not approach those of the other methods, with $R^2 = 0.43$, RMSEP = 7.01, SEP = 6.68, and RPD = 1.2. The complete results for each method are given in Table IV.

TABLE IV

Summary of the coefficient of determination (R^2), root mean square error of prediction (RMSEP), bias, standard error of performance (SEP), ratio of performance to deviation (RPD), and slope of the soil clay content prediction models of partial least squares (PLS), nearest neighbor spectral correction (NNSC), direct standardization (DS), and external parameter orthogonalization (EPO)

Parameter	PLS	NNSC	DS	EPO
R^2	0.69	0.58	0.43	0.63
RMSEP	8.27	5.71	7.01	5.23
Bias	−6.52	0.33	−2.21	−0.06
SEP	5.11	5.73	6.68	5.25
RPD	1.6	1.4	1.2	1.5
Slope	0.52	0.72	0.55	0.61

DISCUSSION

Our attempt to eliminate the moisture effect was examined and compared to two commonly used methods, EPO (Roger *et al.*, 2003) and DS (Ji *et al.*, 2015). The first method predicts clay content directly from the wet spectrum, whereas the second predicts the dry spectrum of a known sample, followed by prediction of its clay content. Whereas the EPO results were comparable to those of our methods, the EPO requires building a new model for each soil property in question. The DS results were far from those of our methods in predicting clay content.

The advantage of the methods suggested in this paper lies in the improved accuracy in predicting clay content (as an example soil attribute) using a soil sample with unknown moisture status. Moreover, the ability to predict the dry soil spectra rather than a specific soil property may enable a full soil property assessment in the future, which cannot be done with the EPO. The soil attribute used herein (clay) can be replaced with any other desired soil attribute, so the suggested methods are more robust. Diminishing the moisture effect could enable normalization of all proxy analyses to the SSL condition, thereby providing a better means of exploiting global SSLs directly, as was done by Castaldi *et al.* (2018) using the LUCAS SSL and APEX data.

The two suggested approaches enabled improving clay content prediction models by simply adding new samples to both calibration and training sets with no need to consider their moisture content. Nonetheless, it should be noted that this paper is only a proof of concept for an idea on how to cope with the moisture effect in soils, paving the way to a more robust method. Accordingly, further work is needed on more soil types under various moisture conditions from different SSLs.

CONCLUSIONS

The goal of this study was twofold. First, we proposed a method for dry soil spectrum prediction in the presence of unknown water content. Next, we showed that the predicted spectrum could be utilized for assessment of clay content and that it provided good estimation in comparison to the use of wet soil spectra. Clay content is an important soil property, and it has been investigated in many studies by VNIR analysis, either by using all of the spectral information or based on the depth of specific clay absorption features at 1 400, 1 900, and 2 200 nm under dry conditions.

In this study, we suggested two independent approaches for executing the task of dry soil spectrum prediction: PLS regression (Wold *et al.*, 2001) and NNSC. Each of these methods creates a single model that can be used to predict the dry spectra of a given soil type. The PLS technique uses the well-established PLS regression, whereas the NNSC method performs wet soil spectral correction by nearest neighbor analysis from a predefined spectral database. It should be noted that both of the constructed methods assume a linear dependence between the dry and wet soil spectra with 30–200 g kg^{-1} water content.

The dry soil spectrum accuracy was measured using SAM and ASDS, which resulted in low prediction errors of less than 8% and 14%, respectively. Later, our hypothesis was tested using the dry soil spectra predicted *via* PLS and NNSC to predict soil clay content, which resulted in $R^2 = 0.69$ and 0.58, respectively. The satisfactory estimations of soil clay content demonstrated the potential of the suggested methods. Although the EPO performed slightly better than the NNSC for predicting the clay content, it did not provide the "dry spectrum", which is important for additional spectral analysis.

To conclude, this research demonstrates the possibility of obtaining dry soil spectra under laboratory conditions and provides impetus for future field experiments. In addition to predicting dry soil spectra accurately, the presented methods have the advantage of being applicable to various soil types using the same model. Therefore, this study paves the way to applying our methods in VNIR imaging spectroscopy to reduce the time and cost of future pedological, agricultural, and environmental analyses.

ACKNOWLEDGEMENTS

The first author Yaron Ogen and the second author Shira Faigenbaum-Golovin contributed equally to this work. The first author Yaron Ogen wishes to thank the Porter School of Environmental Studies, the GEO-CRADLE Project (The European Union's Horizon 2020 Research and Innovation Programme) (No. 690133), the Ministry of National Infrastructures, Energy, and Water Resources of Israel (No. 212-17-025), and the Ministry of Agriculture of Israel (No. 13-21-0002) for financial support and Dr. Gil Eshel from the Ministry of Agriculture and Rural Development of Israel for providing the laser diffraction particle-size analyzer (Mastersizer 2000). The second author Shira Faigenbaum-Golovin wishes to thank the Israel Science Foundation (No. 1457/13) for supporting her research.

REFERENCES

Ackerson J P, Demattê J A M, Morgan C L S. 2015. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. *Geoderma.* **259-260**: 196–204.

Arslan H, Tasan M, Yildirim D, Koksal E S, Cemek B. 2014. Predicting field capacity, wilting point, and the other physical properties of soils using hyperspectral reflectance spectroscopy: Two different statistical approaches. *Environ Monit Assess.* **186**: 5077–5088.

Banin A, Amiel A. 1970. A correlative study of the chemical and physical properties of a group of natural soils of Israel. *Geoderma.* **3**: 185–198.

Ben-Dor E, Banin A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Sci Soc Am J.* **59**: 364–372.

Ben-Dor E, Kindel B, Goetz A F H. 2004. Quality assessment of several methods to recover surface reflectance using synthe-

tic imaging spectroscopy data. *Remote Sens Environ.* **90**: 389–404.

Castaldi F, Chabrillat S, Jones A, Vreys K, Bomans B, van Wesemael B. 2018. Soil organic carbon estimation in croplands by hyperspectral remote APEX data using the LUCAS topsoil database. *Remote Sens.* **10**: 153.

Castaldi F, Palombo A, Pascucci S, Pignatti S, Santini F, Casa R. 2015. Reducing the influence of soil moisture on the estimation of clay from hyperspectral data: a case study using simulated PRISMA data. *Remote Sens.* **7**: 15561–15582.

Chabrillat S, Ben-Dor E, Viscarra Rossel R A, Demattê J A M. 2013. Quantitative soil spectroscopy. *Appl Environ Soil Sci.* **2013**: 616578.

Chabrillat S, Goetz A F H, Krosley L, Olsen H W. 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sens Environ.* **82**: 431–445.

Chang C W, Laird D A, Mausbach M J, Hurburgh Jr C R. 2001. Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. *Soil Sci Soc Am J.* **65**: 480–490.

Christy C D. 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Comput Electron Agric.* **61**: 10–19.

Clark R N. 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. *In* Rencz A N (ed.) Manual of Remote Sensing, Vol. 3, Remote Sensing for the Earth Sciences. John Wiley and Sons, New York. pp. 3–58.

Cohen M J, Prenger J P, DeBusk W F. 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. *J Environ Qual.* **34**: 1422–1434.

Cozzolino D, Morón A. 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil Tillage Res.* **85**: 78–85.

Dunn B W, Batten G D, Beecher H G, Ciavarella S. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis—a case study from the Riverine Plain of southeastern Australia. *Aust J Exp Agric.* **42**: 607–614.

Esbensen K H, Guyot D, Westad F, Houmoller L P. 2002. Multivariate Data Analysis in Practice: An Introduction to Multivariate Data Analysis and Experimental Design. CAMO Process AS, Oslo.

Ge Y F, Morgan C L S, Ackerson J P. 2014. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma.* **221-222**: 61–69.

Islam K, Singh B, McBratney A. 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust J Soil Res.* **41**: 1101–1114.

Janik L J, Forrester S T, Rawson A. 2009. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemom Intell Lab Syst.* **97**: 179–188.

Ji W, Viscarra Rossel R A, Shi Z. 2015. Accounting for the effects of water and the environment on proximally sensed vis-NIR soil spectra and their calibrations. *Eur J Soil Sci.* **66**: 555–565.

Jiang Q H, Chen Y Y, Guo L, Fei T, Qi K. 2016. Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS-NIR spectroscopy. *Remote Sens.* **8**: 755.

Kruse F A, Lefkoff A B, Boardman J W, Heidebrecht K B, Shapiro A T, Barloon P J, Goetz A F H. 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens Environ.* **44**: 145–163.

Liu W D, Baret F, Gu X F, Tong Q X, Zheng L F, Zhang B. 2002. Relating soil surface moisture to reflectance. *Remote Sens Environ.* **81**: 238–246.

Lobell D B, Asner G P. 2002. Moisture effects on soil reflectance. *Soil Sci Soc Am J.* **66**: 722–727.

Minasny B, McBratney A B, Bellon-Maurel V, Roger J M, Gobrecht A, Ferrand L, Joalland S. 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma.* **167-168**: 118–124.

Muller E, Décamps H. 2001. Modeling soil moisture-reflectance. *Remote Sens Environ.* **76**: 173–180.

Nocita M, Stevens A, Noon C, van Wesemael B. 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma.* **199**: 37–42.

Pope R M, Fry E S. 1997. Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. *Appl Opt.* **36**: 8710–8723.

Reeves III J B, Smith D B. 2009. The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Appl Geochem.* **24**: 1472–1481.

Rienzi E A, Mijatovic B, Mueller T G, Matocha C J, Sikora F J, Castrignanò A. 2014. Prediction of soil organic carbon under varying moisture levels using reflectance spectroscopy. *Soil Sci Soc Am J.* **78**: 958–967.

Roberts J J, Cozzolino D. 2016. Wet or dry? The effect of sample characteristics on the determination of soil properties by near infrared spectroscopy. *Trac Trends Anal Chem.* **83**: 25–30.

Roger J M, Chauchard F, Bellon-Maurel V. 2003. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom Intell Lab Syst.* **66**: 191–204.

Roudier P, Hedley C B, Lobsey C R, Viscarra Rossel R A, Leroux C. 2017. Evaluation of two methods to eliminate the effect of water from soil vis-NIR spectra for predictions of organic carbon. *Geoderma.* **296**: 98–107.

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.

Twomey S A, Bohren C F, Mergenthaler J L. 1986. Reflectance and albedo differences between wet and dry surfaces. *Appl Opt.* **25**: 431–437.

Viscarra Rossel R A, Behrens T, Ben-Dor E, Brown D J, Demattê J A M, Shepherd K D, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H, Barthès B G, Bartholomeus H M, Bayer A D, Bernoux M, Böttcher K, Brodský L, Du C W, Chappell A, Fouad Y, Genot V, Gomez C, Grunwald S, Gubler A, Guerrero C, Hedley C B, Knadel M, Morrás H J M, Nocita M, Ramirez-Lopez L, Roudier P, Campos E M R, Sanborn P, Sellitto V M, Sudduth K A, Rawlins B G, Walter C, Winowiecki L A, Hong S Y, Ji W. 2016. A global spectral library to characterize the world's soil. *Earth-Sci Rev.* **155**: 198–230.

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.

Viscarra Rossel R A, Webster R. 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic

database. *Eur J Soil Sci.* **63**: 848–860.

Volkan Bilgili A, 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**: 229–238.

Wijewardane N K, Ge Y F, Morgan C L S. 2016. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma.* **267**: 92–101.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: A basic tool of chemometrics. *Chemom Intell Lab Syst.* **58**: 109–130.

Wu C Y, Jacobson A R, Laba M, Baveye P C. 2009. Alleviating moisture content effects on the visible near-infrared diffuse-reflectance sensing of soils. *Soil Sci.* **174**: 456–465.

Zhu Y D, Weindorf D C, Chakraborty S, Haggard B, Johnson S, Bakr N. 2010. Characterizing surface soil water with field portable diffuse reflectance spectroscopy. *J Hydrol.* **391**: 133–140.

Zornoza R, Guerrero C, Mataix-Solera J, Scow K M, Arcenegui V, Mataix-Beneyto J. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem.* **40**: 1923–1930.