Pre-processing of soil visible and near infrared spectra taken in laboratory and field conditions to improve the within-field soil organic carbon multivariate calibration

K. Kuśnierek*
Department of Soil Science and Remote Sensing of Soils, Adam Mickiewicz University, Poznań, Poland
kus@amu.edu.pl

Abstract

The influence of a mathematical treatment of soil Vis-NIR spectra on multivariate calibration models was investigated. The multivariate calibration is commonly used to derive quantitative information from the reflectance data about the strength of the interaction between the incident radiant energy and the soil organic carbon (SOC) molecules in the organic matter. Among the tested pre-processing procedures: the linearization and normalization of soil spectra outperformed the spectral filtering (i.e.: smoothing, detrending, differentiation) and dataset centering and scaling techniques. The application of min-max normalization to laboratory and field soil spectra reduced the SOC estimation error from 2.5 g kg^{-1} to 2.1 g kg^{-1} and from 3.1 g kg^{-1} to 2.2 g kg^{-1}, respectively.

Keywords: soil organic carbon, pre-processing, maps, partial least squares regression.

Introduction

The recent research on the laboratory and the in-situ application of the portable Vis-NIR spectrometers for measuring soil properties indicates that the accuracy of the spectral reflectance measurements allows the operational scientific and practical usage of this rapid and non-invasive method. There are several reviews of the multivariate calibrations between soil spectral data and the soil organic carbon (SOC) content, one of the most popular soil constituent to model (Vasques et al., 2008; Stenberg et al., 2010). The Vis-NiR reflectance data reveal the strength of the interaction between the incident radiant energy and the SOC bonded in the molecules of the soil organic matter. Yet, the variety of the organic molecules has a high influence on the results of the spectral modeling. The problem of extracting chemical information from the VisNIR spectra is further complicated by the fact that light is scattered by the particles, leading to sample-to-sample variations in the total path length traveled by the photons before they reach the detector (Thennadil et al., 2005). The list of the factors influencing the spectral measurement expands, when the measurements are taken in the field conditions. These additional factors include: soil moisture, soil roughness and the irradiation atmospheric effects (Ben-Dor et al., 1999).

The pre-processing of Vis-NIR soil spectra from the representative post-glacial field-scale study site is tested here to determine its potential in removing the unwanted scattering effects from the spectra taken in the laboratory from ground soil sample and field reflectance spectra of a bare air-dry soil surface of pre-existing soil roughness, taken from the distance of 1.5m at solar illumination.

Materials and methods

The 55 ha study site is located in western Poland (52°30’48’’N, 16°38’27’’E) within the post-glacial landscape of an undulating ground moraine. 68 sampling points were selected using the soil map and aerial photographs of the study area to collect the samples of the soil upper horizon. The SOC content was determined in laboratory by the Walkley-Black method.

Additionally, the soil texture, reaction, and the carbonates content, were described by the appropriate standard methods used in soil science.

The SOC content was calibrated with the spectral reflectance data using the partial least squares (PLS) regression in the ParLeS software (Viscarra Rossel, 2008). The leave-one-out cross-validation of the model was used here to verify the calibration. The studied pre-processing methods included: filtering (i.e.: smoothing, detrending, differentiation), linearization (i.e.: logarithm transformation), normalization (i.e.: min-max, multiplicative scatter correction) as well as dataset centering and scaling techniques.

The spectral measurements were taken in the VisNIR region (350-2500nm) with ASD FieldSpec3 spectrometer. In the laboratory conditions the 6.5 W halogen lamp within the Mug Probe was used to illuminate a standard laboratory soil sample. In the field, the spectra were collected from a bare, air-dry soil surface of pre-existing soil roughness induced by tillage, illuminated by the solar irradiance at the solar zenith angle between 40° and 45° in September 2009, at a distance of 1.5 m.

Results

The loamy soils of the study site are characterized by: the mean sand to silt to clay ratio of 71% to 21% to 8%, the rather alkaline reaction with the mean carbonates content of 1.5%. In the undulating terrain, the nearly 5% slope gradient, the relative altitude differences up to 4 m, and the variable water regime influenced the formation of a diverse soil cover of Haplic Luvisols and Gleyic Chernozems. The spectra of the representative samples of these soil types are depicted in the Figure 1a, and their transformation by the min-max pre-processing method resulting in the best improvement of the SOC multivariate modeling is shown in the figure 1b. The atmospheric absorption bands needed to be removed from the analysis due to the high noise in the data. The accuracy parameters of the soil organic carbon calibration, using raw and pre-processed spectra from laboratory and field spectral measurements is given in the table 1. To test the practical application of the in-situ spectral reflectance measurements, the maps of the SOC content, and the deviation of the spectra derived maps from the reference method are developed (figure 2).

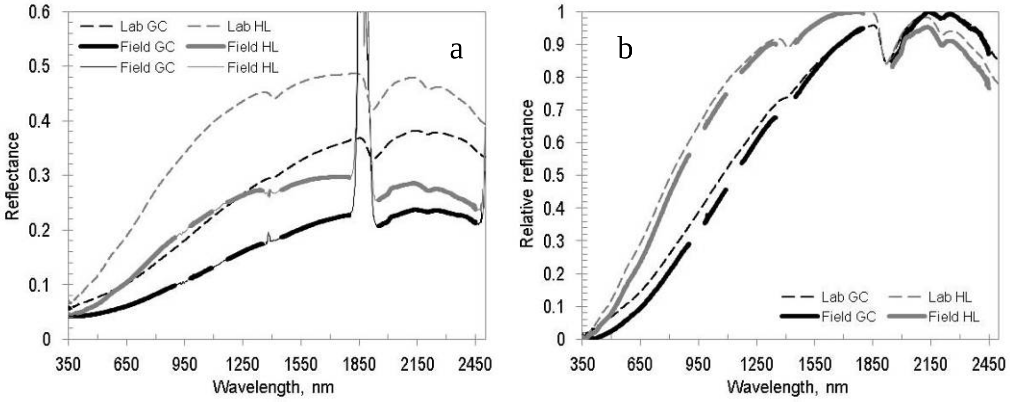

Figure 1. The sample spectra of Gleyic Chernozems (GC, dark lines) and Haplic Luvisols (HL, bright lines) taken in the laboratory (solid) and the field (dashed) conditions (a) and their transformation by the min-max normalization (b).

Table 1. The accuracy parameters of the soil organic carbon multivariate modeling using raw and pre-processed spectra from laboratory and field spectral measurements.

	Laboratory					**Field**				
	n	n.LV[1]	R^2	RMSE ($g\ kg^{-1}$)	RPD	n	n.LV[1]	R^2	RMSE ($g\ kg^{-1}$)	RPD
raw reflectance	**68**	**6**	**0.86**	**2.53**	**2.65**	**68**	**6**	**0.78**	**3.12**	**2.15**
	LINEARIZATION									
absorbance (ABS)	68	6	0.88	2.28	2.94	68	2	0.88	2.27	2.95
Kubelka Munk	68	7	0.89	2.28	2.94	68	3	0.87	2.47	2.71
	NORMALIZATION									
min-max (MMN)	**68**	**8**	**0.9**	**2.1**	**3.19**	**68**	**7**	**0.89**	**2.19**	**3.06**
MSC	68	6	0.89	2.24	2.99	-	-	-	-	-
	FILTERING									
continuum removal	68	6	0.85	2.6	2.58	-	-	-	-	-
Sav.Gol. derivative(SGD)	68	5	0.86	2.47	2.71	-	-	-	-	-
	SCALING									
mean centering	68	3	0.83	2.87	2.33	68	3	0.84	2.75	2.44
median cent. (MEC)	68	8	0.86	2.52	2.66	68	8	0.82	2.89	2.32
	COMBINED METHODS									
ABS_MEC	68	10	0.88	2.33	2.88	68	10	0.86	2.5	2.68
ABS_MMN	68	9	0.89	2.24	2.99	68	7	0.88	2.34	2.86
ABS_MMN_SGD_MEC	68	9	0.9	2.14	3.13	-	-	-	-	-
ABS_MMN_MEC	68	10	0.9	2.13	3.15	68	10	0.81	3.06	2.19

[1] number of latent variables

Discussion

The results confirms that using the selected pre-processing methods before the multivariate calibration of SOC content it is possible to reduce the uninformative variance in the soil spectra caused by the inter-particle light scattering. Here, the SOC calibration accuracy for the laboratory soil spectra increases by 17%. It is also shown that the additional scattering effects present in the field reflectance spectra of a bare air-dry soil surface of pre-existing soil roughness, taken from the 1.5m distance, at solar illumination, may be also considerably reduced by the pre-processing methods. The 30% increase of the modeling accuracy between raw and the min-max normalized spectra (the best performing method for the studied dataset) by the field spectroscopy was observed. The correlation between the performance of the methods in the laboratory and field experiments is indicated here. The pre-processing methods combinations did not yield the expected synergy to further increase the modeling accuracy, and the methods based on moving windows and general trends were not applicable for the field spectra with the gaps after removing the noisy bands.

Conclusions

It has been shown that the use of the pre-processing methods allows reducing the uninformative variance from the soil spectra taken in laboratory and field conditions. The application of the pre-processed field spectra allows producing the soil maps of the accuracy, corresponding to the maps derived from the laboratory spectroscopy. The comparison between the results of this study and the previously published results indicates that the selection of the best performing pre-processing method is dataset dependant. Therefore, the future research on automated unsupervised multivariate modeling procedures using a large number of pre-processing methods is desired.

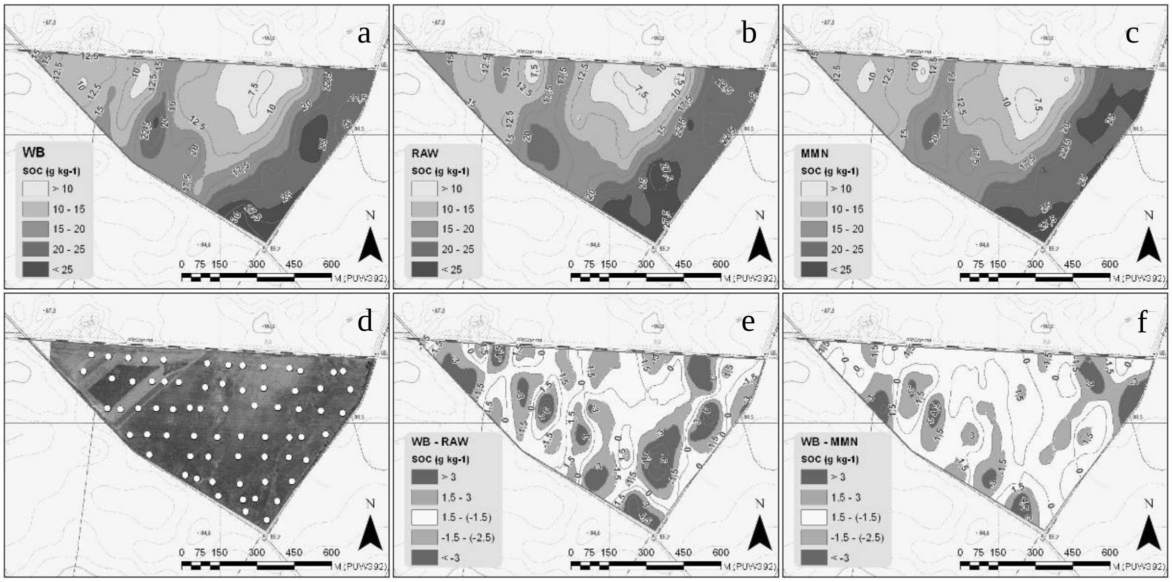

Figure 2. Soil organic carbon content mapped using the reference Walkley-Black method (a), and the multivariate calibration on raw reflectance spectra (b) and min-max pre-processed spectra (c); the sampling points within the test site (d) and the deviation of the spectra derived maps from the reference method for raw (e) and transformed spectra (f).

Acknowledgements

This study was funded by the Polish Ministry of Science and Higher Education (research project no. NN 306 050 739).

References

Ben-Dor, E., Irons, J. R., and Epema, G. F. (1999). Soil reflectance. In: Remote Sensing for the Earth Sciences: Manual of Remote Sensing **3**, edited by N. Rencz, John Wiley & Sons, New York, NY: pp. 111-188.

Stenberg, B., Viscarra Rossel, R., Mounem Mouazen, A., and Wetterlind J. 2010. Visible and near infrared spectroscopy in soil science. In: Advances in Agronomy **107**, edited by D.L. Sparks, Academic Press, Burlington, 2010, pp. 163-215.

Thennadil, S. N., and Martin, E. B. 2005. Empirical preprocessing methods and their impact on NIR calibrations: a simulation study. Journal of Chemometrics **19** 77-89.

Vasques, G., Grunwald, S., and Sickman, J. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* **146** 14-25.

Viscarra Rossel, R. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics and Intelligent Laboratory Systems **90** 72-83.