Development of near infrared spectral library of Danish soils

F. Deng*, M. Knadel, A. Thomsen, M.H. Greve
Dept. of Agroecology and Environment, Faculty of Agricultural Sciences, Aarhus University, Blichers Allé 20, PO Box 50, DK-8830 Tjele
Fan.Deng@agrsci.dk

Abstract

This paper describes the methodology for the establishment of a Danish soil Near InfraRed Spectroscopy (NIRS) library. In order to make an efficient application of NIRS in the field for soil property mapping, it is necessary to establish a NIRS library for global calibration. Representative 3,534 samples from a 7 km grid sampling were chosen to cover variability in the geographical area of Denmark. Partial least square regression was used to build a regression model between SOC and the spectra. The data set was divided into three subsets: calibration, validation and prediction. Outliers were removed from spectral and reference data. The calibration result for non-organic soils was R^2=0.81, RMSE=0.22, while validation was R^2=0.81, RMSE=0.22 and RPD=2.3. This indicates good prediction abilities. The results were also tested on independent datasets, with the following results: R^2=0.82, RMSEP=0.21 and RPD 2.4. Likewise, the calibration result for organic soils was R^2=0.82, RMSE=13.70 while validation was R^2=0.81, RMSE=13.5 with RPD=2.4. Prediction of the independent data set was also satisfactory with the R^2=0.86, RMSEP=13.80 and RPD= 2.7.

Keywords*:* Soil Library, Vis-NIR Spectroscopy, Soil Organic Carbon, Partial Least Square Regression

Introduction

Visible and near-infrared reflectance (VIS-NIR) spectroscopy has developed into a physical non-destructive, rapid and highly reproducible method providing inexpensive prediction of soil physical, chemical and biological properties (Reeves et al., 1999; Rossel et al., 2006). It is relatively easy to use and requires only small quantities of samples for traditional laboratory analysis. Therefore, it provides a powerful tool for proximal soil sensing, digital soil mapping, soil monitoring and process modelling. Soil organic carbon (SOC) is an important factor of soil fertility and plays a critical role in the global carbon cycle (Christopher & Raupach, 2004). The objective of this paper was to develop spectral models to predict SOC content from Danish soils.

Materials and methods

Soil library

The Danish soil library consists of 70,000 samples. The soils are diverse and classified as Alisols, Anthrosols, Arenosols, Cambisols, Fluvisols, Gleysols, Histosols, Leptosols, Luvisols, Phaeozem, Podzols, and Regosols (FAO 1998).The great majority represent sandy soil with low SOC content, however highly organic peat soils are also present.

Within this soil library, 3,330 agriculture soils from 752 profiles collected on a nation-wide 7 km grid were chosen to cover all the variations of Danish soil types. These profiles originated from the field survey in the period from 1986 to 1990. The selected samples consist of 3,288 mineral soils and 42 organic soils. Due to a weak representation of soils high in SOC content, we supplemented with additional 204 organic samples (SOC≥6%) from the Kyoto wetland survey

project, which were collected in 2009 and 2010. Typically, upland fields in Denmark incorporate small patches with elevated SOC content, while wetland fields patches with relatively low SOC. In order to avoid a highly skewed distribution of the data (Stenberg et al., 2010), subdivision of the soil into non-organic (0-12%) and organic (6-55%) was decided for upland and lowland mapping, respectively (Table 1).

Table 1 Minimum, maximum, mean and standard deviation (SD) of SOC in the two sets

	N	Min	Max	Mean	SD
Non-organic	3,330	0	11.2	0.8	1.2
Organic	204	6	55.2	24.5	13.6

Sample preparation and laboratory analysis

The samples were air dried and sieved to 2 mm. Reference organic carbon content was measured by combustion using a LECO CN-2000 instrument (LECO Corp., St. Joseph, Michigan).

Spectra acquisition

Reflectance spectra were collected with the LabSpec5100 instrument (ASD Inc, Boulder, Colorado, USA) covering the visible and near-infrared range between 350 nm and 2,500 nm. Soil samples were homogenized before scanning. They were scanned by using a high intensity mug-light. White reference was taken every five measurements. Two replicates of each sample were taken and averaged automatically by IndioPro 6.0 software (ASD Inc, Boulder, Colorado, USA).

Processing methods

The Unscrambler 10.1 (CAMO, Oslo, Norway) was used for the spectroscopy analysis. Different spectra pre-treatments were tested, including Savitzky-Golay derivatives, Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV) and detrending. Principal Component Analysis (PCA) was introduced for the identification of outliers. Samples with Mahalanobis distance H>3 were considered spectral outliers and removed for further analysis (Shenk and Westerhaus, 1991). Partial Least Square Regression (PLSR) was used to build a regression model for SOC prediction. The non-organic as well as the organic dataset was divided into calibration, validation and prediction subsets. Root Mean Square Error (RMSE) and R^2 were used as criteria for choosing the best regression model. In addition, the ratio of standard error of prediction to standard deviation (RPD) was calculated as RPD=SD/RMSEP, indicating predictive abilities of the model. According to Chang and Laird (2002) and their RPD classification in soil science, RPD>2 indicates a model with good prediction ability, 1.4<RPD<2 is an intermediate model yet to be improved, while RPD<1.4 have no prediction ability. For the non-organic model, due to skewed distribution of SOC content, SOC values were transferred to square root. Wavelengths ranging from 1,000 to 2,500 nm were used as predictors. On the other hand, the SOC content of the organic dataset was normally distributed. Therefore, no transformation was utilized. Spectra ranging from 400 to 2500 nm were used as predictor for building the model.

Results and Discussion

The results of non-organic and organic NIRS models and predictions for the SOC are presented in Table 2. Savitzky-Golay's first derivative with 15 smoothing points appeared to be the best pretreatment technique to build non-organic model. Two, five and eight outliers were removed from the calibration validation and prediction set of the non-organic dataset, respectively. The number of factors used for the calibration was seven. The statistical results from calibration

were R^2=0.81 and RMSE=0.22, while for validation R^2=0.80, RMSE=0.22 and RPD=2.3. The results indicated a good correlation between NIRS and SOC content. The prediction on the third independent dataset confirmed the robustness of the model, with validation R^2=0.82, RMSE=0.21 and RPD=2.4 (Figure1a).

Table 2 Performance of non-organic and organic models

		N	Mean	Min	Max	R^2	SD	RMSE	RPD
Non-organic#	Cal	1,108	0.71	0.00	3.35	0.81	0.51	0.22	
	Val	1,105	0.71	0.00	3.07	0.80	0.50	0.22	2.3
	Pre	1,102	0.70	0.00	3.03	0.82	0.50	0.21	2.4
organic	Cal	131	24.6	6.00	55.20	0.82	13.70	5.77	
	Val	43	24.5	6.50	50.00	0.81	13.50	5.73	2.4
	Pre	22	24.1	6.10	49.00	0.86	13.80	5.12	2.7

SOC data have been square root transformed

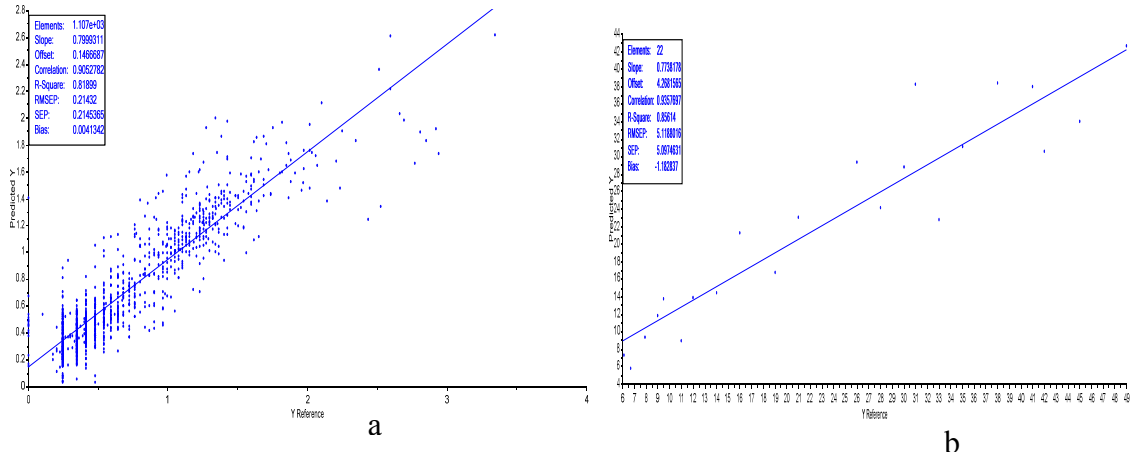

Figure 1. NIRS Calibration, validation and prediction of SOC for non-organic soils

Spectra pretreatment did not improve calibration models for organic dataset. Thus only raw spectra were the predictors in the final analysis with nine factors. Within organic data set two and six outliers were deleted from calibration and validation subsets. The R^2 for calibration and validation was 0.82 and 0.81; RMSE was 5.77 and 5.73, respectively, with RPD value 2.4. The prediction results from this model show relatively low RMSEP value (5.12) in respect to large SOC range and a higher RPD=2.7 (Figure 1b).

Conclusions

The results show that, when dividing the data into non-organic and organic subsets, SOC prediction with NIRS works well in comparison with the traditional laboratory analysis. The two models developed will further be applied for global prediction in relation to proximal soil sensing in Denmark.

Acknowledgements

The authors wish to thank Karin Dyrberg, Stig T. Rasmussen and Bodil Stensgaard, for their support in laboratory analysis, as well as Birte H. Jensen for her corrections in English. We also want to thank two anonymous referees for their comments and suggestions.

References

Chang, C.W., and D.A. Laird. 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science 167:110-116.

Christopher B. Field, and M.R. Raupach. 2004. The global carbon cycle: integrating humans, climate, and the natural world. Island Press. 526 pp.

FAO, 1998. World reference base for soil resources. Food and Agriculture Organization of the United Nations, Rome.

Reeves, J.B., G.W. McCarty, and J.J. Meisinger. 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. Journal of Near Infrared Spectroscopy 7:179-193.

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik, and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75.

Shenk, J.S., Westerhaus, M.O., 1991a. Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. Crop Science 31, 469–474.

Stenberg, B., R.A. Viscarra Rossel, A.M. Mouazen, and J. Wetterlind. 2010. Visible and Near Infrared Spectroscopy in Soil Science. Advances in Agronomy.107:163-215.