Predicting Oxidizable Carbon Content via Visible- and Near-Infrared Diffuse Reflectance Spectroscopy in Soils Heavily Affected by Water Erosion

Radim VAŠÁT, Radka KODEŠOVÁ, Aleš KLEMENT and Ondřej JAKŠÍK

Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic

Abstract

Vašát R., Kodešová R., Klement A., Jakšík O. (2015): Predicting oxidizable carbon content via visible- and near-infrared diffuse reflectance spectroscopy in soils heavily affected by water erosion. Soil & Water Res., 10: 74–77.

Soil spectroscopy represents a low-cost alternative to routine time-consuming and expensive laboratory analyses. Its ability to measure a wide range of different chemical and physical soil properties was shown previously in many studies. Particularly, for organic carbon content, a reliable prediction accuracy is usually achieved. This is due to strong spectral signature of soil organic carbon and other distinct spectral implications of soil characteristics strongly tied to it, e.g. soil colour. All the known studies, however, deal with situation where the study area is fully covered (either in the manner of design- or model-based sampling approach) with calibration points. But in many cases the sampling strategy was initially designed for other purposes, falling outside requirements of spectroscopy for proper model calibration. Hence, here we attempt to test the ability of soil spectroscopy in the situation when only a minor isolated part (the steepest one) of the study area was sampled for calibration points, and predictions were made for its several time larger surroundings. For model training we used Partial Least Squares Regression (PLSR) technique and four different spectra pre-treatment methods (Savitzky-Golay smoothing, first and second derivative, and baseline normalization via continuum removal). Results show high potential ($R^2 \approx 0.70$–0.80) of the method for rough terrain landscapes strongly affected by water erosion, even if the distance from calibration to prediction points is large.

Keywords: Partial Least Squares Regression; spectra pre-treatment; soil assessment

The content of oxidizable carbon (C_{ox}) is an important indicator of the condition and quality of soils. It is routinely determined under laboratory conditions using e.g. the modified Tjurin method (Skjemstad & Baldock 2008) or other standard methods. The laboratory measurements, however, are often costly and time-consuming and hence can be usually done only for a limited number of soil samples. On the other hand, soil spectroscopy has no such constraints, and once the prediction model was properly calibrated it can be used repeatedly for infinite number of samples with much less effort. The ability of soil spectroscopy, especially in visible and near infrared region (400 to 2500 nm), for accurate soil organic carbon prediction was broadly described in scientific literature

(e.g. Gholizadeh *et al.* 2013; Vašát *et al.* 2014). From possible techniques, that are suitable to relate the spectra measurements with oxidizable carbon content, especially Partial Least Square Regression (PLSR) is most frequently used (Visscara Rossel *et al.* 2006). The prediction accuracy may be further increased if spectra pre-treatment methods such as Savitzky-Golay smoothing, first and second order derivative, and baseline normalization via continuum-removal are employed (Gholizadeh *et al.* 2013).

Most studies, apparently, deal with the situation where the study area is covered with calibration points specifically for spectroscopy purposes (either in the manner of design- or model-based sampling approach), so the whole extent of the area is taken

into account. In praxis, however, one often faces the situation when the sampling scheme was initially designed for other purposes falling outside requirements of spectroscopy for proper model calibration. For example, only a minor isolated part of an area might be chosen for calibration sampling, and predictions are required for the rest (several times larger) of the area. This is also the case with this study. Hence, we want to test the ability of soil spectroscopy in the situation where the distance from calibration to prediction points is rather large.

MATERIAL AND METHODS

Study site. The 100 hectare study area located in the south-east of the Czech Republic (South Moravia, Brumovice municipality) can be described as an intensively cultivated piece of arable land with coarse terrain landscape and noticeable water erosion impact. All the soils are developed on loess parent material which is rich in carbonates. Originally the entire area was covered with Haplic Chernozem, degraded mainly at steep parts (the rich humic topsoil horizon was eroded, partially including the parent substrate), while in the lower parts transported humic soil material possibly (depending on terrain conditions) mixed with loess prevailed (Zádorová *et al*. 2013). According to the World Reference Base of Soil Resources (FAO 2014) the main soil units were classified as Haplic Chernozem (top flat parts), Regosol (steep parts), colluvial Chernozem (depressions in

the upper and middle parts), and Colluvial soil (the bottom-most parts). More detailed description of the area conditions can be found in Zádorová *et al*. (2011) and Jakšík *et al*. (2015).

Soil sampling. Originally, the sampling was carried out to study the impact of water erosion on soils and to trace the fate of eroded soil material. Totally 202 sites were visited within this campaign following a judgement sampling design, which may be roughly described as a regular grid with varying spacing (Figure 1a). The goal was to cover different conditions in the field (plateaus, slopes, basins) as evenly as possible. Furthermore, for even more detailed investigation a six-hectare sub-plot in the northwest part of the area was sampled with much higher density (Figure 1b). From this dense sample a subset was further detached to determine some specific, more demanding soil characteristics (Figure 1b, empty circles). As a result we distinguished three subsets to be used as independent calibration and test sets. Subset A of 32 samples (Figure 1b, open circles), subset B of 67 samples (Figure 1b, full circles), and subset C of 107 samples (Figure 1a, cross marks). For C_{ox} analysis only topsoil samples (up to the depth of 20 cm) were used.

Laboratory analysis and spectra collection. Samples were air-dried, ground, and mixed thoroughly before they were sieved to particle fraction ≤ 0.25 mm. C_{ox} was measured using the dichromate redox titration method (Skjemstad & Baldock 2008), when the wet oxidation ($K_2Cr_2O_7$) was fol-
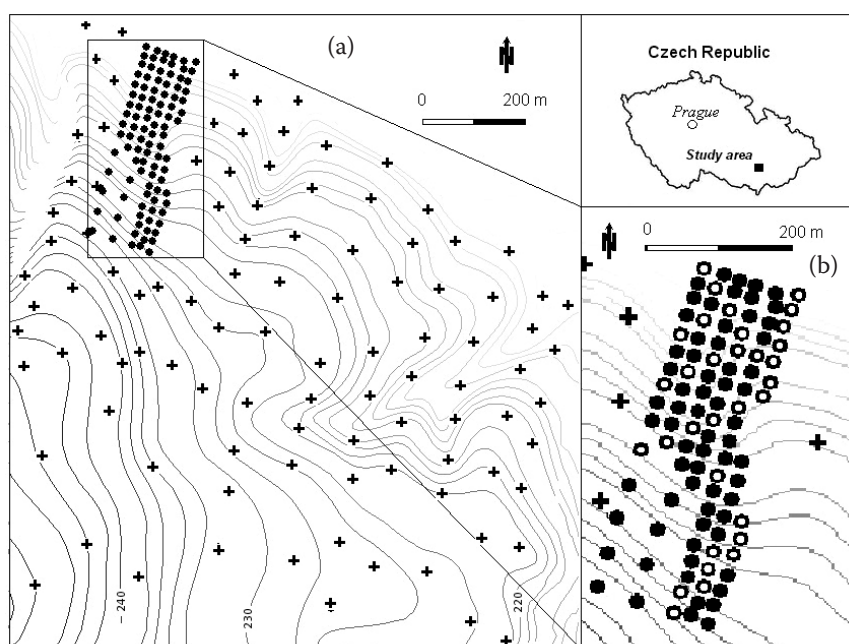

Figure 1. The study area with contours of altitude (graduated grey lines) and sampling designs: (a) subset C (107 samples, cross marks only); (b) subset A (32 samples, open circles) and subset B (67 samples, full circles)

lowed by the potentiometric titration with ferrous ammonium sulphate.

For soil spectra scanning purposes the samples were placed in Petri dishes and the surface of the soil was aligned. Soil spectra were collected for all 202 samples under laboratory conditions using a high intensity contact probe within wavelength range of 350–2500 nm by FieldSpec 3 device (Analytical Spectral Devices Inc., Boulder, USA). The spectral resolution was 3 nm (region 350–1000 nm) and 10 nm (region 1000–2500 nm). The radiometer bandwidth for the region 350–1000 nm was 1.4 nm while it was 2 nm for the region 1000–2500 nm. Calibration of the spectroradiometer was repeatedly done after each of the ten runs by spectralon standard white reference panel (Halon). Finally, the raw spectra were converted into spectral reflectance.

Spectra pre-processing. Four different types of the most commonly used spectra pre-treatments were prepared prior to model calibration. Firstly, the raw spectra were smoothed using the Savitzky-Golay algorithm in order to remove artificial noise which is caused by the spectroradiometer instrument. Secondly, for the smoothed spectra their first and second order derivatives were computed using the Savitzky-Golay algorithm again. And last, the smoothed spectra were subjected to baseline normalization via continuum-removal procedure. All four types of pre-processed spectra were used separately for model calibration and validation to see their effect on prediction accuracy.

PLSR calibration and assessment of prediction accuracy. PLSR models were fitted with pls R package (Mevik & Wehrens 2007) employing the classical orthogonal scores algorithm. Optimal number of PLSR latent variables was determined by minimizing the value of Root Mean Squared Error of Prediction (RMSEP) by leave-one-out cross-validation. Overall prediction accuracy measurements are given, either as results of leave-one-out cross-validation (re-using training set for validation) or independent validation (using external test set for validation), in the meaning of coefficient of determination (R^2) and RMSEP.

We consider two different scenarios: (1) Using subset A as the training set for PLSR calibration, and subsets B and C as the test sets for model validation either separately (subset B only) or together (merged subsets B and C). (2) Using merged subsets A and B as the training set and subset C as the test set.

RESULTS AND DISCUSSION

The results (Table 1) show that the prediction accuracy gets lower, noticeably with the increasing distance from calibration points to prediction points. This is well documented by scenario 1 where the inclusion of more distant points (subset C) to the test set leads to lower prediction accuracy achievement. This decline of prediction accuracy, however, is not dramatic and even so the prediction can still be considered reasonably accurate. Such behaviour may be related to the fact that with increasing distance the soil variability increases too. It was further shown that the cross-validation gives too optimistic overview of the quality of prediction in comparison to the real state (cross-validation offers much better

Table 1. Accuracy measurements for two scenarios and four different spectra pre-treatment methods

	Smoothed spectra only		1st derivative		2nd derivative		Baseline normalization	
Scenario 1								
	R^2_{cv}	RMSEP$_{cv}$	R^2_{cv}	RMSEP$_{cv}$	R^2_{cv}	RMSEP$_{cv}$	R^2_{cv}	RMSEP$_{cv}$
Cross-validation (subset A)	0.93	0.083	0.95	0.068	0.81	0.128	0.87	0.110
	R^2_{ev}	RMSEP$_{ev}$	R^2_{ev}	RMSEP$_{ev}$	R^2_{ev}	RMSEP$_{ev}$	R^2_{ev}	RMSEP$_{ev}$
External validation (subset B)	0.74	0.169	0.75	0.160	0.59	0.220	0.79	0.158
External validation (subsets B+C)	0.69	0.204	0.70	0.201	0.62	0.231	0.71	0.204
Scenario 2								
	R^2_{cv}	RMSEP$_{cv}$	R^2_{cv}	RMSEP$_{cv}$	R^2_{cv}	RMSEP$_{cv}$	R^2_{cv}	RMSEP$_{cv}$
Cross-validation (subsets A+B)	0.85	0.119	0.82	0.130	0.77	0.150	0.86	0.119
	R^2_{ev}	RMSEP$_{ev}$	R^2_{ev}	RMSEP$_{ev}$	R^2_{ev}	RMSEP$_{ev}$	R^2_{ev}	RMSEP$_{ev}$
External validation (subset C)	0.76	0.198	0.74	0.190	0.60	0.241	0.70	0.238

R^2_{cv} – index of determination in cross-validation; R^2_{ev} – index of determination in external validation; RMSEP$_{cv}$ – Root Mean Squared Error of Prediction in cross-validation; RMSEP$_{ev}$ – Root Mean Squared Error of Prediction in external validation

results as compared with external validation). This indicates a high degree of statistical adaptation, which is even more pronounced in scenario 1 where fewer samples were used for calibration.

Concerning the spectra pre-treatment, it is not clear which of the four tested methods has the most positive effect on accuracy of the prediction. The use of Savitzky-Golay smoothing, 1st derivative, and baseline normalization leads to similar accuracy achievements for both scenarios. Only in case of 2nd derivative transformation the predictions accuracy decreased rapidly. Hence, no clear positive effect of tested spectra pre-treatment on prediction quality was observed.

In overall, the accuracies (from external validation) vary from reliable ($R^2_{cv} \geq 0.75$) to rather less reliable ($0.50 \leq R^2_{cv} < 0.75$), which is entirely consistent with other results described in scientific literature (Gholizadeh *et al.* 2014). As such the prediction models can be used for direct measurement of C_{ox} content, as well as a valuable source of auxiliary data for high resolution mapping or screening purposes.

CONCLUSION

The study shows that soil visible- and near-infrared diffuse reflectance spectroscopy (400 to 2500 nm) is a suitable tool for estimating the content of C_{ox} of arable land heavily impacted by water erosion. Using a proper calibration set one may achieve a fairly reliable prediction which is suitable for direct measurement of C_{ox}. The prediction accuracy, however, decreases noticeably when the distance from calibration points to prediction increases. But despite this, even if the distance between calibration and prediction points is large, the results can still be used as a valuable source of auxiliary information for high resolution mapping or screening purposes.

References

FAO (2014): World Reference Base for Soil Resources. Rome, FAO.

Gholizadeh A., Borůvka L., Saberioon M.M., Vašát R. (2013): Visible, near-infrared and mid-infrared spectroscopy application for soil assessment with emphasis to soil organic matter content and quality: State-of-the-art and key issues. Applied Spectroscopy, 67: 1349–1362.

Jakšík O., Kodešová R., Kubiš A., Stehlíková I., Drábek O., Kapička A. (2015): Soil aggregate stability within morphologically diverse areas. Catena, 127: 287–299.

Mevik B.H., Wehrens R. (2007): The pls package: principal component and partial least squares regression in R. Journal of Statistical Software, 18: 1–24.

Skjemstad J., Baldock J.A. (2008): Total and organic carbon. In: Carter M. (ed.): Soil Sampling and Methods of Analysis. 2nd Ed., Boca Raton, Soil Science Society of Canada, CRC Press: 225–238.

Vašát R., Kodešová R., Borůvka L., Klement A., Jakšík O., Gholizadeh A. (2014): Consideration of peak parameters derived from continuum-removed spectra to predict extractable nutrients in soils with visible and near-infrared diffuse reflectance spectroscopy (VNIR-DRS). Geoderma, 232–234: 208–218.

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.

Zádorová T., Penížek V., Šefrna L., Rohošková M., Borůvka L. (2011): Spatial delineation of OC-rich Colluvial soils in Chernozem regions by Terrain analysis and fuzzy classification. Catena, 85: 22–33.

Zádorová T., Penížek V., Šefrna L., Drábek O., Mihaljevič M., Volf Š., Chuman T. (2013): Identification of Neolithic to Modern erosion-sedimentation phases using geochemical approach in a loess covered sub-catchment of South Moracia, Czech Republic. Geoderma, 195–196: 56–69.

Corresponding author:

Ing. Radim Vašát, Česká zemědělská univerzita, Fakulta agrobiologie, potravinových a přírodních zdrojů, katedra pedologie a ochrany půd, Kamýcká 129, 165 21 Praha 6-Suchdol, Česká republika; e-mail: vasat@af.czu.cz