Polish Journal of Agronomy
2010, 3, 3–9

Visible and near-infrared spectrophotometer for soil analysis: preliminary results

[1]Guillaume Debaene, [1]Jacek Niedźwiecki, and [2]Alicja Pecio

[1]Department of Soil Science Erosion and Land Conservation, [2]Department of Plant Nutrition and Fertilization
Institute of Soil Science and Plant Cultivation – National Research Institute
ul. Czartoryskich 8, 24-100 Puławy, Poland

Abstract. Precision farming, a fast growing technique, is based on within-field variability at the field or farm scale and demands some information about soil physical and chemical properties. In this paper, we present the use of visible and near infrared reflection spectroscopy (VIS-NIRS) in the 400–2200 nm spectral range to predict soil acidity, available Mg, P, and K content, soil organic carbon (SOC) content, and soil clay fraction (<0.002 mm) content, and then present an on-the-go spectrophotometer for in situ measurements of reflectance spectra. Some optimistic preliminary results have been obtained for the prediction of SOC, clay, available Mg, and K content with r^2 (predicted vs. measured values) varying from 0.64 to 0.69. Results also emphasise the importance of the calibration scheme.

key words: precision agriculture, visible and near infrared spectroscopy (VIS-NIRS), soil organic carbon (SOC), partial least square regression (PLS), on-the-go spectrophotometer

INTRODUCTION

Management of agricultural systems is dependent on demands to control the cost of production and to increase productivity. To obtain a better response from inputs in agriculture, numerous analyses are needed so that those inputs can be applied where they best fill their purpose (van Vuuren et al., 2006). VIS-NIRS technology has the potential to detect fine-scale spatial variability of soil. Furthermore, results can be very accurate as Viscarra Rossel and McBratney (2008) have shown in their review. Unfortunately, transition from standard soil testing approach to the adoption of VIS-NIRS method for precision agriculture need a shift of mentality as underlined by van Vuuren et al. (2006).

VIS-NIRS has been used in agriculture for assessing grain, fertilisers and soil qualities (Ben-Dor and Banin, 1995; Faraji et al., 2004; Mohan et al., 2005) and has proven to be a rapid, convenient means of analysing many soil constituents at the same time. Soil properties that have been calibrated with VIS-NIRS include the determination of soil moisture, SOC content, electrical conductivity (EC), cation exchange capacity (CEC), soil acidity, some macro- and microelements (Dunn et al., 2002; Velasquez et al., 2005).

Absorption in the near-infrared spectral region (780–2500 nm) is dominated by molecules that contain strong bonds between light atoms. Specifically, these are molecules that contain C-H, N-H or O-H bonds. This makes the near infrared region particularly useful for measuring forms of carbon, nitrogen and water. VIS-NIRS is a rapid and non-destructive analytical technique that correlates diffusely reflected near-infrared radiation with the chemical and physical properties of materials (Chang and Laird, 2002). One interesting advantage of VIS-NIRS is that the size of spectrometers is rather small so that they can be field-portable (Christy, 2008).

The objective of this work was to investigate the usefulness of VIS-NIRS in determining various soil chemical property (SOC content, soil acidity, content of available Mg, K, and P) and a single physical properties (clay content) in topsoil (0–25 cm) from a soil sampling grid field. With that aim in mind, two calibration schemes have been elaborated. Results discrepancies between the two calibration schemes are discussed in relation to predicted sample localisation and soil texture variability.

MATERIAL AND METHODS

Sample collection and preparation

One hundred and twenty soil samples were collected from a soil sampling grid (Figure 1) in the experimental station of the Institute of Soil Science and Plant Cultivation

---

Corresponding author:
Guillaume Debaene
e-mail: gdebaene@iung.pulawy.pl
tel. +48 81 8863421 ext. 396

Received 3 August 2010

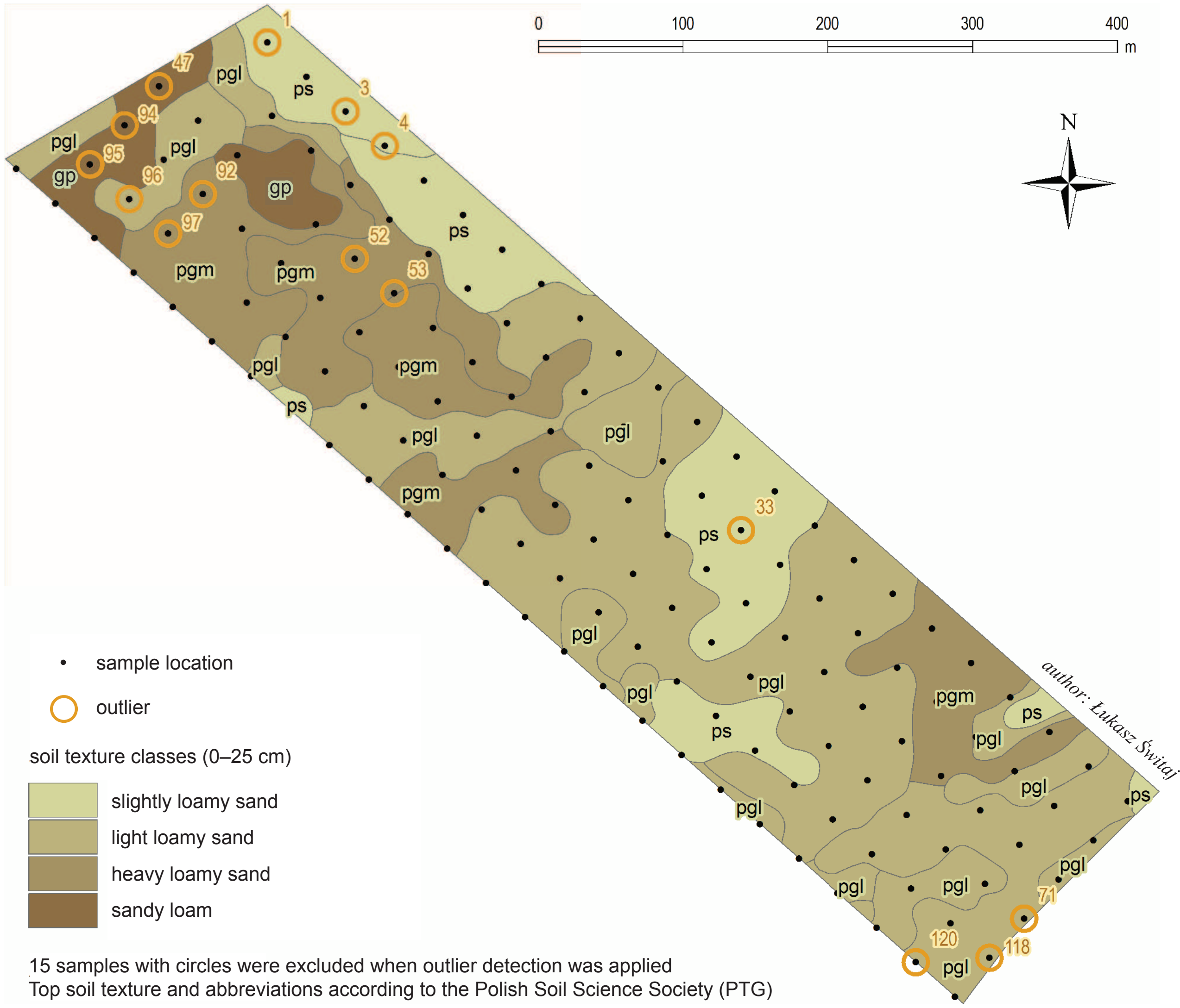


Fig. 1. Map illustrating the Baborówko field and sample locations

(IUNG) in Baborówko near Poznań (Wielkopolska, Poland). The grid belongs to field n°1 the area of which is 9 ha. The station conducts its field trials on an area of 53,6 ha (3 fields) in the following crop rotation: winter wheat, spring barley and winter rape according to the rules of precision agriculture. IUNG uniform agrotechnics recommendations for specific crop technologies and integrated agriculture rules are applied. The soil type at this station is a podzoluvisol (World Reference Base for Soil Resources – WRB). A considerable topsoil texture mosaic is observed on the whole research area but also on the field n°1 in question. The soil grid is georeferenced by means of a global positioning system. Coordinates of the grid centre are N 52°35’1.6”, E 16°38’50.47”. Samples were collected in the topsoil (0–25 cm), air dried and no further treatments were applied in order to be able to compare results with on-the-go measures in the future.

Chemical analyses

Chemical analyses of the samples were performed by the IUNG Main Chemical Laboratory in Puławy. The soil acidity values were measured potentiometrically in 1M KCl solution. Soil organic carbon content was determined by the Tiurin method (K dichromate digestion). The contents of available K, P, and Mg (mg 100 g^{-1} of soil) were determined in calcium lactate-extractable K, calcium lactate-extractable P, and calcium chloride-extractable Mg. The Mastersizer 2000 apparatus with Hydro MU attachment from the Malvern Company was used for the laser diffraction method for particle size distribution. The Mastersizer 2000 is designed for standard determination of grain size distribution of particles within the size range of 0.002–2 mm. It makes use of laser light scattered on measured particles and converts it into particle size distribution. Measurements were made in three replications (a new portion of air-dry soil poured into the measuring system being treated as a replication). Calculations of particle size distribution were carried out using the Fraunhofer and Mie theories (in the case of Mie theory the refractive index of 1.57 and absorption index of zero were used). The laser light wavelength in the apparatus was 466 nm for blue and 633 nm for red light. Measurements (understood as averaging of 30 000 images of laser light diffraction recorded by the detectors) lasted 60 s (30 s for blue and 30 s for red light) and were carried out directly one by one.

System overview

The samples were scanned from 400 to 2220 nm with the Veris VIS-NIR spectrophotometer in bench top mode (Veris Technologies, Salina, KS, USA) by means of the Veris Spectrophotometer Software V1.69. The sample holder is placed against the face of the sapphire window of the shank module that contains a tungsten halogen lamp and fibre optics for transmission to the spectrometer. Samples were scanned 20 times and averaged by the software and data were collected every 5.5 nm of the electromagnetic spectrum. The absorbance of the scanned sample is given by the relation log (1/R) where R is the reflectance.

On-the-go machine

The Veris spectrophotometer acquires absorbance measurements of soil while being pulled through the field. It is built in a shank mounted on a toolbar, and pulled behind a tractor. The shank is lowered into the ground to approximately 7 cm and pulled through the soil at 6 to 10 km h^{-1}. The device makes measurements through a sapphire window mounted on the bottom of the shank. The texture of the sapphire keeps it clean through its journey into the soil. The device is described at length in Christy (2008). The main feature of this spectrophotometer is the possibility to perform real-time measurements via NIR spectroscopy. Another interesting feature is that all spectra pretreatments are realised on the field i.e. (1) data extraction, (2) filtering, (3) clustering. Outliers are removed during the filtering process using Mahalanobis distance. The aim of these 3 steps is to determine the best soil sampling location. According to spectral properties and principal component analysis, with the help of a fuzzy c-means algorithm, clusters of observations are realised. There are as many samples locations as clusters. Each cluster is representative of the overall spectral variation (Naes, 1987). The number of clusters decided by the operator. The sample location is computed to be close to the centre of the cluster. These samples are needed to create VIS-NIRS calibrations for quantitative predictions. The last treatment is the interpolation using Gaussian weighting to average spectra near a sampled location. Spectra near the location are weighted higher than spectra far away. Then the NIR calibration and validation methods are to be chosen by the operator and laboratory computed.

VIS-NIRS calibration

The spectrum from each sample was matched with laboratory analysis data to create a database for calibrations. A multivariate calibration model is required to obtain some practical information from the VIS-NIRS spectra. The Partial Least Square Regression (PLS), a popular multivariate calibration technique for quantitative analysis of NIR spectral data, was used to determine the best correlation between the chemical data and spectra data. The PLS is a dimension reduction technique that seeks a set of latent variables by maximizing the covariance of two variable blocks (i.e., spectra X and concentration Y). Data were calibrated and tested using the R software – Version 2.11.1 (R Development Core Team, Vienna) and the R *pls* package from Mevik and Wehrens (2007). The *pls* package implements a leave-one-out cross-validation on the calibration set. Ten components are taken into account for the PLS regression of the calibration models. In leave-one-out cross-validation each sample is omitted and predicted using a calibration made from the remaining samples. The number of components to use for validation is given by the root mean squared error of prediction (RMSEP) of the calibration set. The calibration models were then used to predict the chemical parameters of the sample prediction sets.

Two sets of calibration were prepared without outlier treatment and two sets with outlier detection based on a new method from Filzmoser et al. (2005). The method is automated to identify outliers in multivariate space and to distinguish between extreme values of a normal distribution and values originating from a different distribution.

RESULTS

Three samples were not analysed for all chemical properties and therefore were removed from the sample database leaving 117 samples. Sample names have not been changed so that they were easier to position on the map. Mean, standard deviation and ranges for each chemical and physical parameters are shown in Table 1. The SOC values displayed a narrow range 0.71–1.45% conforming with typical content levels of Polish soils. The mean clay content is generally lower than the usual content of Polish soils.

Ten randomly selected soil sample spectra are presented in Figure 2. As can be seen, the spectral bands largely

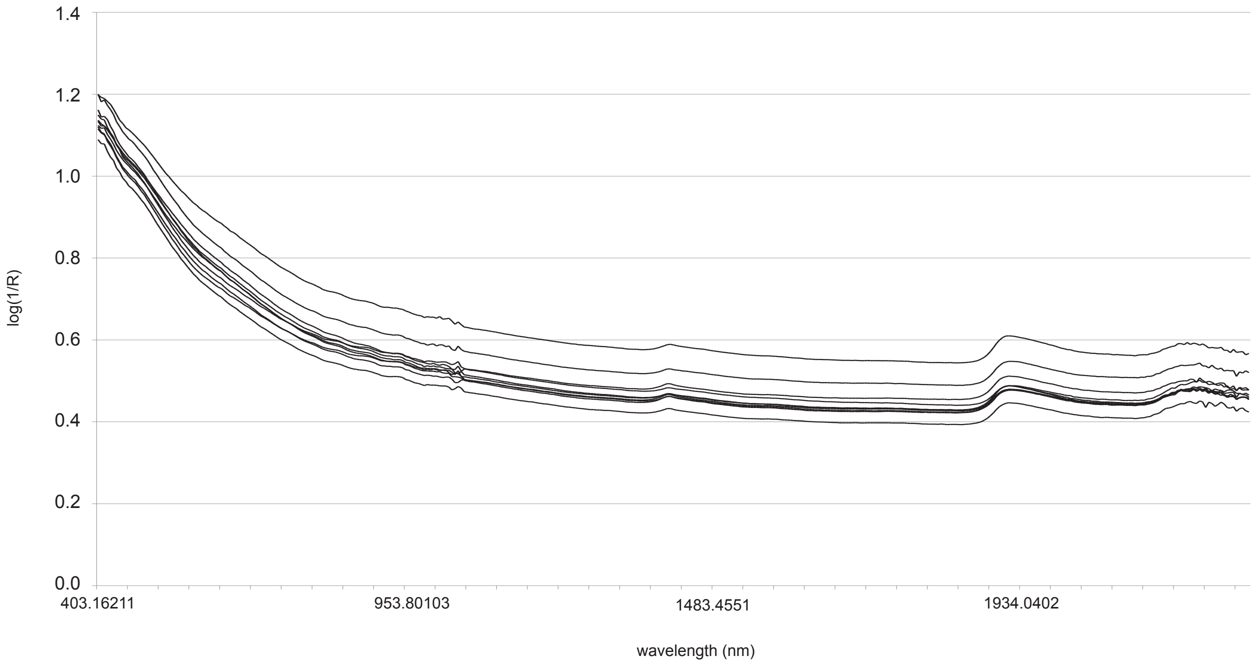
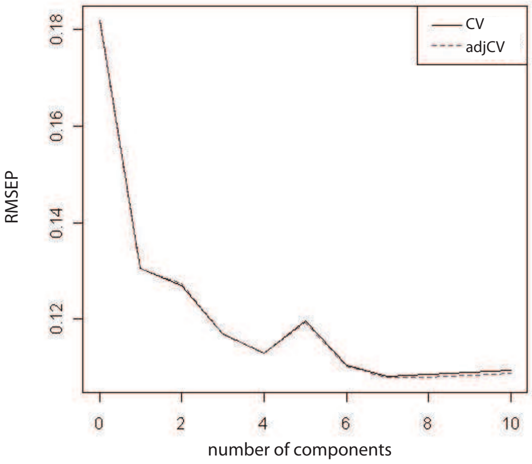


Fig. 2. Near-infrared spectra of 10 randomly selected soil samples.

Table 1. Chemical and physical characteristics of soil samples used in this study.

Variable	Mean ± SD	Range	Mean ± SD	Range
	Series 1–2		Series 3–4	
SOC	1.11 ± 0.20	0.71–1.80	1.08 ± 0.16	0.71–1.45
pH	6.22 ± 0.52	4.99–7.57	6.17 ± 0.47	5.00–7.48
Mg	6.63 ± 2.50	3.10–15.9	6.29 ± 1.97	3.10–12.4
K	13.91 ± 4.95	5.10–44.5	13.07 ± 3.68	5.10–23.4
P	14.40 ± 5.74	5.70–33.6	13.64 ± 4.92	5.70–26.2
clay	1.99 ± 0.82	0.10–3.67	1.95–0.80	0.10–3.38

117 samples were used without outlier detection: series 1 and 2
102 samples left after outlier detection: series 3 and 4
SD – standard deviation
pH – soil acidity
SOC – soil organic carbon content (% soil)
Mg, K, P – content of available Mg, K, P (mg 100 g^{-1} of soil)
clay – clay content (% soil < 0.002 mm)



CV – cross-validation estimate
adjCV – bias-corrected cross-validation estimate
RMSEP – root mean squared error of prediction

Fig. 3. Cross-validated RMSEP curves for the soil organic carbon content: series 1.

overlap. The raw (log 1/R) spectra of all soils have similar shapes with the strongest absorption in the visible range and small peaks (e.g. 1400 and 1900 nm) in the near infrared region. One can deduce from this that spectral exploitation is only possible with the help of advanced multivariate statistics. Three to five components for the validation of the different calibration sets were generally sufficient. As can be seen in Figure 3 which presents as an example the estimated RMSEP as functions of the number of components for the SOC content of the first series, four components seem to be enough.

Without outlier treatment

The first calibration was undertaken by taking the samples from the series as they come in sequence from the grid. The samples were divided into two sets. The first 90 samples were used for calibration and the last 27 samples of the series were used for the validation of the calibration. The second calibration scheme was designed as in Dunn et al. (2002). Starting from sample n°1, data of every fourth sample were moved to a separate file for use as the validation set (30 samples). In that case, samples to be predicted are distributed evenly on the whole area of the field. Coefficient of determination r^2 and root mean squared error of prediction (RMSEP) of the two calibration procedures without outlier detection are given in Table 2 and 3. RMSEP is an indicator of how close validation standards – standards that are not in the calibration – are to the calibration line. There are two results of prediction, the leave-one-out (r_1^2, $RMSEP_1$) for the cross-validation of the calibration set and the validation for the prediction set (r_2^2, $RMSEP_2$). It is to notice that sometimes for a small dataset, only the cross-validation r_1^2 and $RMSEP_1$ are used to test the prediction ability of a model.

The best predictions (r_2^2) are achieved for soil organic carbon content (SOC) and for the content of available Mg for both the two calibration schemes. The first scheme giving somewhat better predicted values. The calibration scheme of series 2 shown better predictive ability for available K and clay content than series 1.

It can be noticed that the prediction obtained for the second series with the sample validation set evenly scattered in the field (calibration scheme 2) are overall better than for the first series where the sample validation set focused on the western border of the field (calibration scheme 1). With the first scheme, four of six properties were very poorly predicted or not at all for the content of available P, but presented reasonable cross-validated prediction r_1^2 (Table 2).

With outlier treatment

Because in large databases outliers can be easily unnoticed, we decided to check for outliers using the Filzmoser et al. (2005) method. All outliers detected were removed (circles in Figure 1) on the base of their chemical properties. Calibration schemes are the same as above. It gives for scheme 1 a calibration set of 85 samples with a validation set containing 17 samples (series 3) and for scheme 2 a calibration set of 78 samples with a validation set of 24 samples. Table 4 and 5 present the r^2 and RMSEP for the two calibration with outliers detection. The prediction of series 3 without outliers are worse than the prediction with all samples, especially the regression coefficient of the validation sets (r_2^2). All chemical properties but SOC were not or very poorly predicted. It is to notice that the r_1^2 are much better than the r_2^2 and sometimes even satisfactory (Mg or SOC content). The calibration scheme 2 of the fourth series of samples predicted the SOC and available K content. All other cross-validated and predicted properties were unsatisfactory (r^2 0.30–0.48).

Overall, the calibration scheme 2 is better working with or without outliers.

Table 2. Calibration and prediction results for soil properties: series 1.

Variable	r_1^2	$RMSEP_1$	r_2^2	$RMSEP_2$
SOC	0.61	0.11	0.65	0.14
pH	0.55	0.36	0.24	0.49
Mg	0.71	1.14	0.69	2.16
K	0.51	3.63	0.29	4.15
P	0.41	4.61	0.01	5.42
clay	0.61	0.50	0.30	0.58

pH – soil acidity
SOC – soil organic carbon content (% soil)
Mg, K, P – content of available Mg, K, P (mg 100 g^{-1} of soil)
clay – clay content (% soil < 0.002 mm)
r_1^2 – calibration set regression, r_2^2 – validation set regression,
RMSEP – root mean squared error of prediction

Table 3. Calibration and prediction results for soil properties: series 2.

Variable	r_1^2	$RMSEP_1$	r_2^2	$RMSEP_2$
SOC	0.64	0.12	0.50	0.14
pH	0.38	0.39	0.40	0.46
Mg	0.69	1.39	0.51	1.50
K	0.42	3.96	0.61	3.05
P	0.28	4.75	0.40	5.10
clay	0.60	0.53	0.64	0.48

Explanations – see Table 2

Table 4. Calibration and prediction results for soil properties: without outlier, series 3.

Variable	r_1^2	$RMSEP_1$	r_2^2	$RMSEP_2$
SOC	0.60	0.10	0.56	0.13
pH	0.47	0.35	0.02	0.46
Mg	0.69	1.02	0.38	2.39
K	0.57	2.43	0.31	2.98
P	0.41	3.76	0.07	5.24
clay	0.51	0.55	0.02	0.96

Explanations – see Table 2

Table 5. Calibration and prediction results for soil properties: without outlier, series 4.

Variable	r_1^2	$RMSEP_1$	r_2^2	$RMSEP_2$
SOC	0.48	0.11	0.68	0.11
pH	0.39	0.36	0.30	0.40
Mg	0.48	1.34	0.48	1.63
K	0.45	2.49	0.68	2.64
P	0.34	3.96	0.32	4.36
clay	0.36	0.41	0.48	0.60

Explanations – see Table 2

DISCUSSION

Results have pointed out that the type of calibration and the use of outlier detection have an important impact on the validity of prediction of the chemical and physical properties of soil samples with NIRS technology. Differences between results of the two calibration can be explained as follows. The validation set of the first scheme was composed of samples situated on the edge of the studied area, and not included in the calibrated area. On the other hand, the validation samples of the second scheme were spread on the whole area and therefore included in the calibration area. In the light of our results, and even if regression coefficients are not absolutely satisfying, it appears that evenly scattered sample datasets are better predicted than sample grouped on the edge of the studied area. This seems logical because of the better representation of chemical and physical properties in both calibration and validation sets for scattered samples.

The first scheme with all samples has failed to predict most of the soil properties but SOC and Mg content. In addition to the fact that predicted samples are outside the calibration area, as said above, samples 94 to 98, from a region of high topsoil texture variability present higher content of clay, available P and K but also Mg. Those five samples have a great impact on the regression result. Removing them improves the prediction. In the future more importance should be given to this region of complicated mosaic topsoil texture. More samples are needed to be sampled in this part of the field to better represent the overall variability of topsoil characteristics and then to improve the prediction ability of the calibration model.

The third and fourth series of calibration were obtained with an outlier treatment. Fifteen samples were removed from the dataset. As can be seen in Figure 1, most of outliers are situated on the border and in the northern part of the field. This can be explained by the fact that other field trials are conducted on neighbour fields and some contamination may occur. Moreover, some of these outliers are located in depleted region (tractor tracks) where sometimes water accumulates and in an area with sandy loam and slightly loamy sand. The effect of water and clay on the other soil properties are that some characteristics are unusually high and are detected as outlier by the algorithm probably without actually being so. As explained above, samples 94 to 98 are somewhat outside the range of values of the soil characteristics. This is another clue as to sample more this part of the field and it shows that choosing sampling location is an essential task in VIS-NIRS.

It seems, in the light of the prediction results that scheme 1 without outliers poorly predicts soil characteristics but for SOC (Table 4). Nevertheless, some properties seems to be predicted with the leave-one-out cross-validation. However, as stated by Dardenne et al., (2000) there is always an over estimation of cross-validation when no validation sets are taken into account. Doing a cross-validation with the calibration and validation set together even improves the cross-validation prediction. The reason for the poorly predicted results can be the fact that there are not enough samples in the calibration set and that removing the outliers is narrowing too much the range of the chemical and physical characteristics, as regression is generally improved when data are widespread.

Scheme 2 greatly improved the prediction of all properties with outlier detection.

Overall, the results of the present study are preliminary and can be improved. One hundred and twenty samples seemed to be enough for that work since most studies reviewed by Viscarra Rossel and McBratney (2008) included 100 to 200 samples. Some of the samples which seem to be far of the regression line will soon be re-analysed, scanned once more and then reinserted in the calibration model. The authors have tried to add several more samples to the models (but unfortunately from another neighbouring field) with a great improvement of the prediction results except for available P. The lack of prediction of available P can be related to the fact that the variable is poorly distributed in comparison with other variables presented in this work. The same reason was assumed by Dunn et al. (2002) for exchangeable Na. In some cases, removing few samples could greatly improve the accuracy of the prediction. Other outlier removing strategies need also to be checked. A data pre-treatment as the one from Chang et al. (2001) increase prediction regression for some properties (e.g. $r^2 = 0.82$ for SOC). Nevertheless, our results for prediction of SOC, clay, available Mg and K content are similar to that of Chang et al. (2001) and of Dunn et al. (2002) even if their studies were based on samples from a much wider area. This work is also a first step towards the analysis *in situ* with the on-the-go VIS-NIRS spectrophotometer. Some maps of chemical properties and especially of the carbon content will be realised and compared with on-the-go maps.

Results presented above are preliminary in nature, 300 more samples from the soil grid (3 fields) will be analysed later, added to the calibration, and compared with the results that will be measured in situ with the on-the-go spectrophotometer. With this aim in mind, the on-the-go spectrophotometer for in situ measurements of reflectance spectra is also described in this paper.

CONCLUSIONS

In the view of our results, near-infrared reflectance spectrometry is a simple to implement, non-destructive method that could be used to predict some soil properties. Results of this study have pointed out that:

1. VIS-NIRS is a suitable method for estimating SOC, clay, available Mg and K content;

2. The calibration scheme is very important and calibration samples should encompass the whole spectrum of chemical and physical characteristics of the soil;

3. More attention has to be paid to areas with important soil texture variability;

4. Outlier detection is not always the best way to obtain a good prediction.

REFERENCES

Ben-Dor E., Banin A., 1995. Near-Infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am., 59: 364-72.

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

Chang C.W., Laird D.A., Mausbach M.J., Hurburgh C.R., 2001. Near-Infrared Reflectance Spectroscopy-Principal Component 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.

Dardenne P., Sinnaeve G., BaetenV., 2000. Multivariate calibration and chemometrics for near infrared spectroscopy: which method? J. Near Infrared Spec., 8: 229-237.

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

Faraji H., Crowe T., Besant R., Sokhansanj S., Wood H., 2004. Prediction of moisture content of potash fertilizer using NIR spectroscopy. Biosyst. Eng., 46: 3.45-3.48.

Filzmoser P., Garrett R.G., Reimann C., 2005. Multivariate outlier detection in exploration geochemistry. Comput. Geosci., 31: 579-587.

Mevik B-H., WehrensR., 2007. The *pls* Package: Principal Component and Partial Least Squares Regression in R.J. Stat. Software, 18: 1-24.

Mohan L.A., Karunakaran C., Jayas D.S., White N.D.G., 2005. Classification of bulk cereals using visible and NIR reflectance characteristics. Canad. Biosyst. Eng., 7: 7.7-7.14.

Naes T., 1987. The design of calibration in near infrared reflectance analysis by clustering. J. Chemometr., 1: 121-134.

Van Vuuren J.A.J., Meyer J.H. Claassens A.S., 2006. Potential Use of Near Infrared Reflectance Monitoring in Precision Agriculture. Commun. Soil Sci. Plant Anal., 37: 2171-2184.

Velasquez E., Lavelle P., Barrios E., Joffre R., Revers F., 2005. Evaluating soil quality in tropical agroecosystems of Colombia using NIRS. Soil Biol. Biochem., 37: 889-898.

Viscarra Rossel R.A., McBratney A.B., 2008. Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping. 165-72. Hartemink A.E., McBratney A.B., Mendonça-Santos L., Digital soil mapping with limited data. Developments in Soil Science series. Elsevier Science, Amsterdam.