Comparison of the Performance of Two Vis-NIR Spectrometers in the Prediction of Various Soil Properties

Marie-Christine Marmette[1], Viacheslav Adamchuk[1], Jacques Nault[2], Salman Tabatabai[3], Robert Cocciardi[4].

[1]Department of Bioresource Engineering, McGill University, Sainte-Anne-de-Bellevue, Québec, Canada.
[2]Logiag, Châteauguay, Québec, Canada.
[3]Department of Agroecology, Aahrus University, Tjele, Denmark.
[4]Malvern Panalytical Ltd., Saint-Laurent, Québec, Canada.

Abstract. *Spectroscopy has shown capabilities of predicting certain soil properties. Hence, it is a promising avenue to complement traditional wet chemistry analysis that is costly and time-consuming. This study focuses on the comparison of two Vis-NIR instruments of different resolution to assess the effect of the resolution on the ability of an instrument to predict various soil properties. In this study, 798 air dried and compressed soil samples representing different agro-climatic conditions across Québec (Canada) were analyzed using Vis-NIR spectroscopy. Vis-NIR spectra of all soil samples were collected using a laboratory setup of a field spectrometer operating in the range from 350 - 2200 nm (P4000, Veris Technologies, Salina, Kansas, USA) and the ASD FieldSpec® 4 Standard-Res Spectroradiometer (Malvern Panalytical Ltd, Malvern, United Kingdom) operating from 350 - 2500 nm. In addition to the analytical techniques, successful prediction of soil properties depends on sensor calibration. In this research, three spectral pre-processing methods were compared (standard normal variate, first and second derivatives, all with a Savitzky-Golay filter), the results were produced using partial least squares regression (PLSR) and the models were selected according to the R^2 of a 15-fold cross-validation. The results of each combination of soil property (extractable P, K, Ca, Mg, Al, SOM and CEC), data calibration method and instrument were assessed in terms of RMSE of the prediction and the R^2 for the linear regression between measured and predicted values. FieldSpec gave better predictions for K (R^2 = 0.34, RMSE = 145 kg/ha), Al (R^2 = 0.60, RMSE = 164 ppm), SOM (R^2 = 0.69, RMSE = 0.97%) and CEC (R^2 = 0.62, RMSE = 2.94 cmolc/kg) and Veris gave better predictions for P (R^2 = 0.11, RMSE = 142 kg/ha), Ca (R^2 = 0.63, RMSE = 1260 kg/ha) and Mg (R^2 = 0.58, RMSE = 232 kg/ha). It was not possible to conclude which instrument performs better.*

Keywords. *Precision agriculture, proximal soil sensing, soil spectroscopy, Vis-NIR*

Introduction

Soil analysis is at the heart of precision agriculture applications and there is a need to find alternatives to, or complementary methods for, traditional laboratory analysis. Spectroscopy is a promising avenue for soil analysis that has the advantage of being rapid, cost-effective and versatile. Logiag (Châteauguay, Québec, Canada) is an innovative agronomic support company providing services to growers and agribusinesses across Eastern Canada and Northeastern USA. Through cooperation with the National Research Council (NRC, Boucherville, Québec, Canada), they have developed LaserAg technology, a novel method to analyze soil properties using air-dried and compressed soil samples employing Laser-Induced Breakdown Spectroscopy (LIBS). This technology allows a significant reduction of soil laboratory analysis costs and improves logistics. Although the application of LIBS to measure soil properties has been evaluated by a number of researchers (Viscarra Rossel *et al.*, 2011), the technological cycle implemented by Logiag has been adopted for commercial use and was certified through ISO 17025. At the same time, other technology developers are engaged in diffuse soil reflectance (Viscarra Rossel *et al.*, 2016) and machine vision. To this day, there is no reported analysis on the complementarity of LIBS and other spectral analytics performed on the same set of soil samples; this is essential to employ sensor fusion concepts and enhance the reliability of soil test results. This results in two research questions: 1. Are there alternative technologies that have compatible analytical capabilities? 2. Can it be determined that adding less challenging spectral measurement techniques to LIBS will increase the accuracy of soil properties assessment? It is known that sensor fusion allows for the removal of bias for indirect data inference, which is typical to chemometrics methods (Adamchuk *et al.*, 2011).

The three main regions of the electromagnetic spectrum that are used for soil analysis are the visible (Vis, 350 - 700 nm), the near-infrared (NIR, 700 – 2 500 nm) and mid-infrared (MIR, 2 500 - 25 000 nm). Vis spectroscopy involves energy absorption through electronic transitions. NIR and MID electromagnetic waves are too low in energy to induce electronic transitions. MID spectroscopy studies fundamental vibrations of covalent bonds whereas NIR involves overtones and combinations of those vibrations. Various qualities of instruments exist for each spectroscopic method. In addition to the analytical technique, successful prediction of soil properties depends on sensor calibration. The main methods employed in soil spectroscopy are partial least squares regression (PLSR), support vector machines (SVM), random forest (RF), multivariate adaptive regression splines (MARS), and regression trees (CART).

The ultimate goal of this research was to assess the compatibility of LaserAg with other technologies to improve the process. This is done in three major steps: measurement of soil spectral characteristics using instruments of varying quality and spectral ranges (color, Vis-NIR and MIR), analysis and comparison of different data processing techniques (PLSR, SVM, RF, MARS and CART), and dissemination of the results obtained and development of the plan for implementation of the research findings in the Logiag business structure. The present study focuses on the comparison of the performance of two Vis-NIR spectrometers for the prediction of extractable phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), aluminum (Al), soil organic matter (SOM) and cationic exchange capacity (CEC) using PLSR for the calibration. More specifically, the objectives of this paper are to (i) determine which pre-processing among standard normal variate, first derivative and second derivative works the best for every combination of soil property and instrument, (ii) build models with PLSR to predict seven soil properties (P, K, Ca, Mg, Al, OM and CEC) and (iii) determine which instrument performs the best predictions based on the root mean squared error (RMSE).

Materials and Methods

Experimental Data

Soil Samples and Reference Data

The 798 soil samples used in this study were collected on various farms throughout the province of Quebec, Canada. The samples were analyzed in two different laboratories to provide the reference values of extractable phosphorous (P), potassium (K), calcium (Ca), magnesium (Mg), aluminum (Al), soil organic matter (SOM) as well as cation exchange capacity (CEC): EnvironeX Group (Québec, Québec, Canada) and GEOSOL Laboratory (Synagri, Saint-Hyacinthe, Québec, Canada). Table 1 presents the properties analyzed, the methods employed, and the number of samples for each laboratory.

Table 1 - Laboratory methods of soil analyses

Soil properties	GEOSOL Laboratory 401 samples		Environex Group 397 samples	
	Method	Units	Method	Units
P	Mehlich III with plasma	kg ha^{-1}	Mehlich III with plasma	kg ha^{-1}
K	Mehlich III with plasma	kg ha^{-1}	Mehlich III with plasma	kg ha^{-1}
Mg	Mehlich III with plasma	kg ha^{-1}	Mehlich III with plasma	kg ha^{-1}
Ca	Mehlich III with plasma	kg ha^{-1}	Mehlich III with plasma	kg ha^{-1}
Al	Mehlich III with plasma	ppm	Mehlich III with plasma	ppm
SOM	Wackley−Black	%	Loss on ignition	%
CEC	Calculated based on K, Mg and Ca values	meq 100g^{-1}	Calculated based on K, Kg, Ca and buffer pH values.	cmol$_c$ kg^{-1}

The dataset had high variability and all the distributions were positively skewed. Table 2 shows the distribution parameters of each soil property. Reference values over 1200 kg/ha for P, 1500 kg/ha for K, 15 000 kg/ha for Ca, 3000 kg/ha for Mg, 2000 ppm for Al, 15 % for SOM and 45 cmol$_c$/kg for CEC were considered as outliers and discarded from the sample set for the processing.

Table 2 – Distribution parameters of the reference soil properties

Parameter	P (kg ha^{-1})	K (kg ha^{-1})	Ca (kg ha^{-1})	Mg (kg ha^{-1})	Al (ppm)	SOM (%)	CEC (cmol$_c$ kg^{-1})
Minimum	12	40	285	15	188	0.60	7.2
Maximum	1660	2050	17588	3635	2169	54	61
Mean	212	302	4858	535	1111	5.0	19
Median	144	259	4746	391	1085	4.2	18
Standard Deviation	225	218	2249	444	293	4.2	5.8
Skewness	3.0	2.4	0.72	1.7	0.36	7.2	1.4
Number of values	791	797	797	795	798	798	798

To eliminate the bias due to moisture content and bulk density, the samples were air dried, placed in individual plastic cups resistant to high-pressure loads (diameter of 4.2 cm) and compressed under a force of approximately 20 t (196 kN), resulting in 35 MPa pressure. Logiag (Châteauguay, Quebec) acquired the soil spectra with the LIBS method, which left 24 burns concentrated in the middle of each sample.

Spectral Scanning

Among seven different spectral instruments, two systems were used to measure diffuse light reflectance in the Vis-NIR part of the electromagnetic spectrum: Veris® P4000 hydraulic probe (Veris Technologies Inc., Salina, Kansas, USA), hereinafter called Veris, and ASD FieldSpec® 4 Standard-Res Spectroradiometer (Malvern Panalytical Ltd, Malvern, United Kingdom), hereinafter

called FieldSpec. Veris has a spectral range of 342 - 2220 nm with 384 spectral bands and an 8 nm resolution. Veris is designed for field measurements, but for the present application it was installed indoors for laboratory measurements. The probe is 102 cm long and is equipped with a sapphire window and fiber optics. Two detectors acquire the spectrum: Toshiba TCD1304AP Linear CCD Array covering 342 - 1023 nm with 128 bands and InGaAs Linear image sensor G9206-02 covering 2070 - 2220 nm in 256 bands. The light source is a halogen bulb. To minimize the instrument noise, each spectrum was recorded as an average of 30 scans and the calibration was performed on every 20 samples using Avian Reflectance Standards (Avian Technologies LLC, New London, New Hampshire, USA). The other spectrometer used, the FieldSpec, has a spectral range of 350 - 2500 nm and a spectral resolution of 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. Three elements are used to complete the spectrum of 2151 narrow bands: 512 pixels silicon array (350 - 1000 nm) and two Graded Index InGaAs Photodiode for 1001 - 1800 nm and 1801 - 2500 nm. The FieldSpec Contact Probe was used to acquire the spectra. It is equipped with a halogen light bulb, a sapphire window, and optical fiber to transfer the signal to the spectrometer. Each spectrum recorded was the average of 50 scans and a calibration was done every 7 minutes with a Spectralon® panel (Labsphere Inc, North Sutton, New Hampshire, USA).

Spectra Pre-processing

Spectral pre-processing techniques are used to reduce the un-modeled variability in the data and to reduce noise and enhance the features sought in the spectra (Gholizadeh *et al*, 2015; Rinnan *et al*, 2009; Buddenbaum & Steffens, 2012). There is not a single good avenue when it comes to pre-processing; the latter depends on the dataset (Stenberg *et al*, 2010). Since applying the wrong type of pre-processing or applying too severe ones can remove important and valuable information, multiple methods were applied to the dataset using the prospectr R package and the one giving the best results was selected for each soil property and spectrometer combination. The Savitzky-Golay (SG) (Savitzky & Golay, 1964) filter was present in all pre-processing. SG filter is a ubiquitous smoothing method allowing noise reduction (Rinnan *et al*, 2009) that fits a least squares polynomial to a number of data points. Using more data points in the filter window increases the smoothing whereas using higher-degree polynomial as the fitting function decreases the smoothing. A window of 11 bands and a second-order polynomial were used for each pre-processing method since it those settings showed good results for Gholizadeh *et al* (2015), Hong *et al* (2017) and Rinnan *et al* (2009).

The first method employed was the SG filter followed by a Standard Normal Variate (SNV) (Barnes *et al*, 1989). SNV is a scatter correction method that aims to reduce the physical variability between samples due to multiplicative interferences of light scatter and particle size by centering and scaling each spectrum individually:

$$x_f = \frac{x_i - a_0}{a_1} \tag{1}$$

where x_i and x_f are the original and the corrected spectra, a_0 is the average value of the sample spectrum to be corrected and a_1 is its standard deviation. The second pre-processing method used was the First Derivative (FD) and it was done via SG smoothing. Finally, a Second Derivative (SD) was also done via SG smoothing.

Chemometric Analysis

Partial Least Squares Regression (PLSR)

PLSR is a method that relates two data matrices, **X** of predictors and **Y** of responses, by a linear multivariate model. PLSR is used in spectroscopy because it can analyze data with strongly collinear, noisy and numerous **X**-variables. PLSR is close to Principal Component Regression (PCR). Unlike PCR, PLSR models the structure of **Y** and integrates compression and regression steps to select the successive orthogonal factors that maximize the covariance between **X** and **Y** (Wold *et al*, 1983; Wold, 2001). The number of factors used in the model was determined by

cross-validation (CV). The chemometric analysis was done using the pls R package.

Model Development

Selecting the calibration samples based on their spectra is a good way to capture as much as possible of the variation in the larger dataset (Wetterlind *et al*, 2013). The set of 798 samples was separated into training (70%) and test (30%) sets using Kennard-Stone sampling on the **X** matrix, with Mahalanobis distance to describe the **X**-space and three principal components. This method allows the creation of a training set that has a flat distribution over the spectral space (Kennard & Stone, 1969). Since all samples have 3 replicates (except for 18 samples in the Veris set that has only 2 replicates), and since all replicates of a sample must remain in the same set, the Kennard-Stone sampling was done on the average of the sample spectra.

Once the averaging and the sampling were done, a PLSR was run on the non-averaged pre-processed spectra. Using an excessive number of factors contributes to overfit the model and describing noise in the data (Buddenbaum & Steffens, 2012), thus, a maximum of 10 PLS latent factors wes used in the regressions. The averaged coefficient of determination (R^2) of the 15-fold CV wes used to select the number of PLS latent variables:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{2}$$

where y_i is the measured soil property, $\hat{y}_i$ is the predicted soil property, $\bar{y}$ is the mean of measured soil properties and n is the number of spectra used in the validation. To ensure the independence of the calibration and validation sets, all replicates of a sample were kept in the same fold. For each soil property and instrument, a parsimonious and a more complex model were selected if possible. The number of latent variables for each model was selected visually with the break in the graph.

Model Performance Assessment

Once the optimal pre-processing and the number of factors were selected for the simple and complex models, the test sets were used to perform the predictions. The performance of the predictions was assessed with the adjusted coefficient of determination (R^2_{adj}) and RMSE:

$$R^2_{adj} = 1 - \frac{N-1}{N-k}(1-R^2) \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{4}$$

where *N* is the number of spectra and *k* is the number of predictors.

Results and Discussion

PLSR Model Optimization

Once the 15-fold CV was practiced for each model, the CV R^2 according to the number of factors was plotted for the three pre-processing techniques, for each soil property and for each instrument (Figure 1). For FieldSpec, except for the K and CEC (at two factors) where SNV gave the best R^2, FD gave a better R^2 for all properties, followed by SNV and SD that arrived last. For Veris, there was no pre-processing that was generally better than the others, it depended on the soil property and the number of factors. The 15-fold CV showed better R^2 with FieldSpec, especially with K, Al and SOM. The smallest CV R^2 was obtained with P and K and were under 0.4, except for K with FieldSpec beyond 7 factors. The highest CV R^2 was obtained with FieldSpec for CEC where it goes over 0.7 at 5 factors.
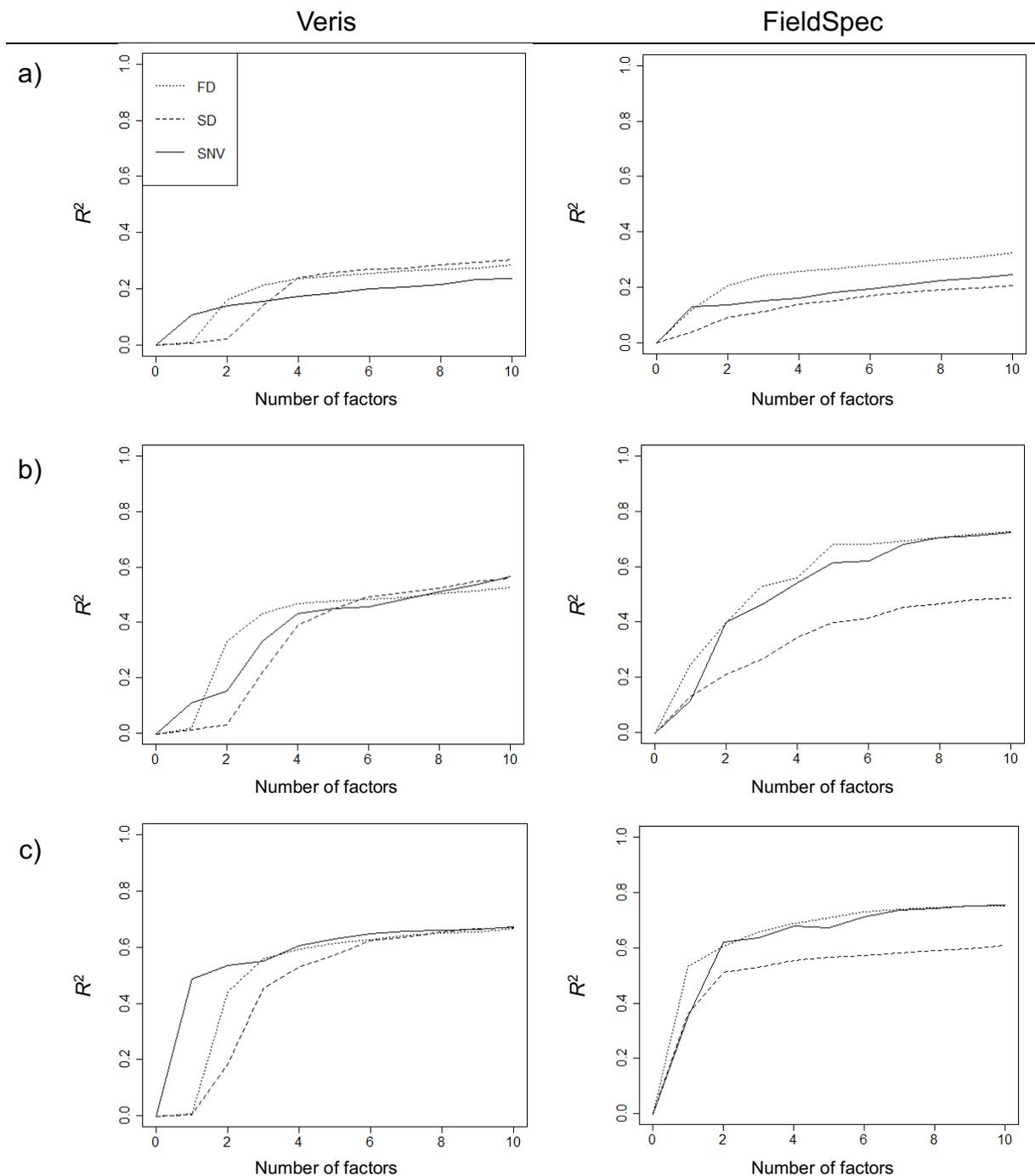
| | Veris | FieldSpec |

Figure 1. Coefficients of determination (R^2) of the partial least squares regression's 15-fold cross-validations with three pre-processing techniques: first derivative (FD), second derivative (SD) and standard normal variate (SNV). Veris and FieldSpec results are presented for (a) extractable phosphorus, (b) soil organic matter and (c) cation exchange capacity.

Predictability of Soil Properties

For every property and every instrument, a parsimonious (S) and a more complex (C) model in terms of the number of PLS factors were selected for the predictions. Table 3 shows the pre-processing methods and the number of factors selected for those two models, as well as the R^2 and RMSE of the CV and the R^2_{adj} and RMSE of the prediction. The best models for every property, marked in bold, were the ones with the lowest RMSE.

Table 3 – Results of predictions

Soil Property	Model	Veris						FieldSpec					
		PP	N	R^2 CV	RMSE CV	R^2_{adj}	RMSE	PP	N	R^2 CV	RMSE CV	R^2_{adj}	RMSE
P (kg ha⁻¹)	S	**FD**	**3**	**0.21**	**181**	**0.11**	**142**	FD	3	0.25	160	0.24	172
	C	SD	4	0.24	173	0.23	144	FD	4	0.26	159	0.26	169
K (kg ha⁻¹)	S	FD	3	0.21	174	0.2	171	SNV	2	0.25	175	0.19	161
	C	-	-	-	-	-	-	**SNV**	**7**	**0.4**	**156**	**0.34**	**1445**
Ca (kg ha⁻¹)	S	FD	3	0.45	1642	0.43	1576	FD	4	0.58	1470	0.53	1297
	C	**FD**	**7**	**0.57**	**1441**	**0.63**	**1260**	-	-	-	-	-	-
Mg (kg ha⁻¹)	S	FD	3	0.59	280	0.56	276	FD	2	0.63	275	0.52	260
	C	**FD**	**5**	**0.66**	**266**	**0.58**	**232**	-	-	-	-	-	-
Al (ppm)	S	SNV	5	0.38	226	0.41	221	FD	3	0.59	191	0.52	179
	C	FD	6	0.58	189	0.54	184	**FD**	**5**	**0.66**	**174**	**0.6**	**164**
SOM (%)	S	FD	4	0.47	1.51	0.52	1.4	FD	3	0.52	1.48	0.56	1.15
	C	SNV	6	0.46	1.6	0.55	1.19	**FD**	**5**	**0.67**	**1.22**	**0.69**	**0.97**
CEC (cmol_c kg⁻¹)	S	SNV	2	0.53	3.97	0.42	3.76	SNV	2	0.62	3.64	0.4	3.46
	C	FD	3	0.56	3.71	0.63	3.36	**FD**	**4**	**0.69**	**3.27**	**0.62**	**2.94**

Results of the predictions according to the model (S or C), the pre-processing technique (PP) and the number of PLSR factors (N). R^2 CV and RMSE CV are the cross-validation's coefficient of determination and root mean squared error. R^2_{adj} and RMSE the prediction's adjusted coefficient of determination and root mean squared error. The model presenting the lowest RMSE for each soil property is in bold.

Veris gave better RMSE results for P, Ca and Mg, whereas better results were obtained with FieldSpec for K, Al, OM and CEC. For three factors with P, Veris obtained an RMSE of 142 kg/ha and an R^2_{adj} = 0.11 using a FD pre-processing whereas FieldSpec gave a better R^2_{adj} of 0.26 but a smaller higher RMSE=169 kg/ha also with FD. This is under results previously obtained by other researchers with R^2_{adj} of 0.48 and 0.73 (Maleki *et al*, 2006; Wetterlind *et al*, 2010), but is higher than the R^2_{adj} = 0.01 obtained by Viscarra Rossel *et al* (2006). The lowest RMSE for P was obtained with the model that had the highest CV RMSE (181 kg/ha) and the lowest CV R^2 (0.21). For K, FieldSpec gave the best RMSE=145 kg/ha with SVN and 7 factors. With an R^2_{adj} = 0.34, this result for K is comparable to the 0.33 and 0.47 previously obtained by Janik *et al* (1998) and Viscarra Rossel (2006). K and P are the only soil properties that had an R^2 below 0.40. For Ca, Veris gave the best prediction (RMSE=1260 kg/ha, R^2_{adj}=0.63) with 7 factors and FD, which is under the R^2 found in the literature that vary between 0.67 and 0.90 (Chang *et al* 2001; Cozzolino & Moron, 2003; Islam *et al*, 2004). Mg also obtained a better prediction with Veris with RMSE = 232 kg/ha and R^2_{adj} = 0.58 (5 factors, FD), which is again under the R^2 varying from 0.59-0.90 found in the literature (Chang *et al* 2001; Cozzolino & Moron, 2003; Islam *et al*, 2004). FieldSpec gave better predictions for Al, SOM and CEC, obtaining respective RMSEs of 164 ppm (5 factors, FD), 0.97 % (5 factors, FD) and 2.94 cmol_c/kg (4 factors, FD). The Al R^2_{adj} of 0.60 is similar to previous results obtained by Viscarra Rossel (2006) for the NIR (0.61) and Janik *et al* (1998) (0.64), but under Buddenbaum and Steffens (2012) results (0.73-0.76). For the SOM, the R^2_{adj} of 0.69 is in the range of previous findings varying from 0.49-0.94 (Nawar *et al*, 2016; Kodaira *et al*, 2009, Wetterlind *et al*. 2010, Kodaira and Shibusawa, 2013). Finally, the R^2_{adj} = 0.62 obtained for the CEC is also in the range of previous findings varying between 0.13 and 0.89 (Viscarra *et al*, 2006; D'Acqui *et al*, 2010; Kodaira & Shibusawa, 2013). Figure 2 shows the measured versus the predicted P, SOM and CEC values of simple and complex models for both Veris and FieldSpec.
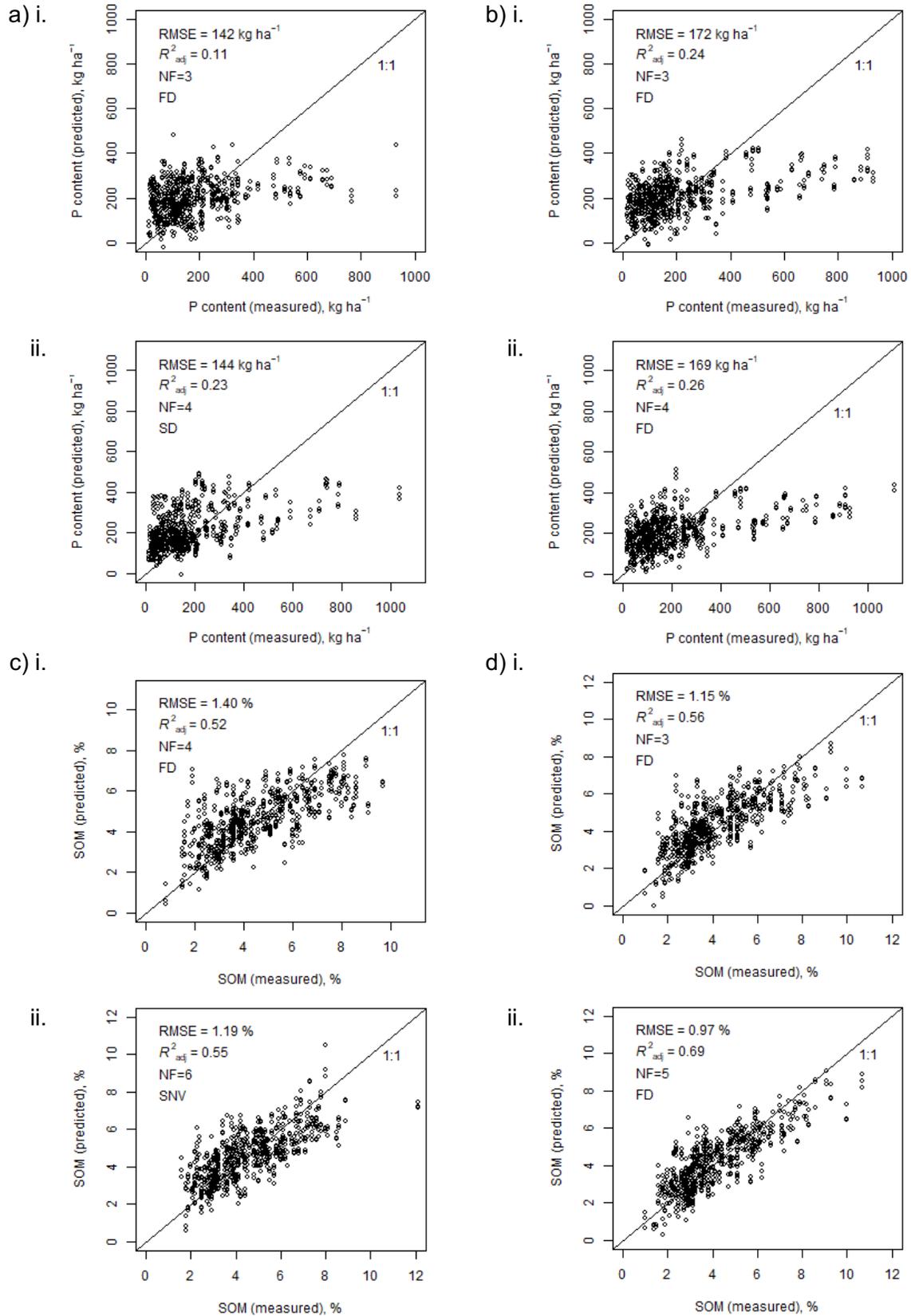
Figure 2. Model testing for Veris and FieldSpec. i. shows the simpler model (S) and ii. shows the more complex model (C). a), c) and e) show the P, SOM and CEC prediction results for Veris. b), d) and f) show the P, SOM and CEC prediction results for FieldSpec.
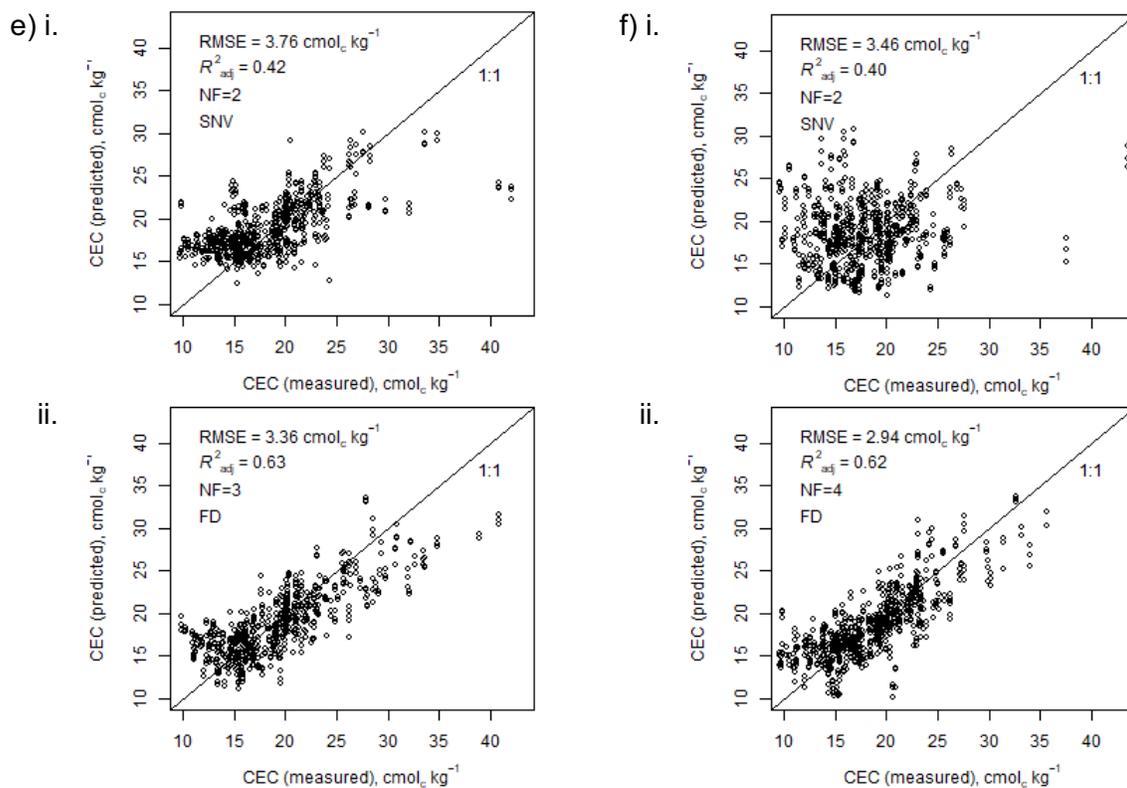
Figure 2 - Continued

Except for the particular case of P, the simplest models (S) never gave better results than the ones including more factors (C). The best model selected between Veris and FieldSpec was the one including more factors for K (7 with FieldSpec vs 3 with Veris), Ca (7 with Veris vs 4 with FieldSpec), Mg (5 with Veris vs 2 with FieldSpec) and CEC (4 with FieldSpec vs 3 with Veris). Also, the R^2 of CV continued to increase beyond the number of factors selected. Therefore, a higher number of factors could have been used in the regressions and could have given better results. Indeed, while the present research used between 4 and 7 factors, numerous researchers used between 10 and 30 factors to predict soil properties in the Vis-NIR region (Reeves *et al*, 1999; Reeves & McCarty, 2001; Walvoort & McBratney, 2001; McCarty *et al*, 2002; Viscarra *et al*, 2006; Viscarra *et al*, 2009; Vohland *et al*, 2014).

Conclusions

The performances of two spectrometers in the Vis-NIR range were compared in this study for the prediction of seven soil properties. FieldSpec gave better results for four of the properties (K, Al, SOM, CEC) and Veris for the three other properties (P, Ca, Mg). FD was the best pre-processing method for six of the properties (P, Ca, Mg, Al, OM, CEC) and SNV for K. Even if FieldSpec gave better predictions for four out of seven properties, it is not possible to state that it is better than Veris since the number of factors seemed to have more influence on the quality of the predictions than the instrument itself. Further research should be conducted to properly compare the performance of FieldSpec and Veris. In that research, the spectrometer sensors junctions could be excluded of **X** and a variable selection technique, such as Variable Importance in Projection, could be used to retain only the significant wavelength in **X** (Mehmood *et al*, 2012; Farres *et al*, 2015). Also, the method to determine the number of PLSR factors could be done using the minimum RMSE obtained through the CV (Viscarra *et al*, 2006; Vohland *et al*, 2011; Kodaira & Shibusawa, 2013; Nawar *et al*, 2016; Hong *et al*, 2017). Finally, a combination of pre-processing methods might be tested in different sequences to see if the combination of normalizations and derivatives could provide better results.

Acknowledgements

References

Adamchuk, V.I., Viscarra Rossel, R.A., Sudduth, K.A., & Schulze Lammers, P. (2011). Sensor fusion for precision agriculture. In: *Sensor Fusion – Foundation and Applications*, Chapter 2, 27-40, C. Thomas, ed. Rijeka, Croatia: InTech.

Barnes, R. J., Dhanoa, M. S., Susan J., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772-777.

Buddenbaum, H., & Steffens, M. (2012). The effects of spectral pretreatments on chemometric analyses of soil profiles using laboratory imaging spectroscopy. *Applied and Environmental Soil Science*, 1–12.

Chang, C. W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65(2), 480-490.

Chong, I., & Jun, C. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78, 103-112.

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

D'Acqui, L.P., Pucci, A., & Janik, L.J. (2010). Soil properties prediction of western Mediterranean islands with similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy. *European Journal of Soil Science*, 61(6), 865–876.

Farres, M., Platikanov, S., Tsakovski, S., & Tauler, R. (2015). Comparison of variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics*, 29, 528-536.

Gholizadeg, A., Boruvka, L., Saberioon, M. M., Kozak, J., Vasat, R., & Nemecek, K. (2015). Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectra features. *Soil & Water Resources*, 10(4), 218-227.

Hong, Y., Yu, L., Chen, Y., Liu, Y., Liu, Y. , Liu, Y., & Cheng, H. (2017). Prediction of Soil Organic Matter by VIS–NIR Spectroscopy Using Normalized Soil Moisture Index as a Proxy of Soil Moisture. *Remote Sensing*, 10(28), 1-17.

Islam, K., Singh, B., Schwenke, G., & McBratney, A. (2004). Evaluation of vertosol soil fertility using ultra-violet, visible and near-infrared reflectance spectroscopy. SuperSoil 2004: Third Australian New Zealand Soils Conference, 5–9 December 2004, University of Sydney, Australia.

Janik, L.J., Merry, R.H., & Skjemstad, J.O. (1998). Can mid infra-red diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture*, 38(7), 681– 696.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11, 137–148.

Kodaira, M., Shibusawa, S., Ninomiya, K., & Kato, Y. (2009). Farm mapping techniques for effective soil management in large-scale farming. *JSAI Journal*, 18(3), 110-121.

Kodaira, M., & Shibusawa, S. (2013). Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. *Geoderma*, 199, 64-79.

Maleki, M.R., Van Holm, L., Ramon, H., Merckx, R., De Baerdemaeker, J., & Mouazen, A.M. (2006). Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. *Biosystems Engineering*, 95 (3), 425–436.

McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., & Kimble, J.M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. *Soil Science Society of America Journal*, 66, 640-646.

Malvern Panalytical. ASD FieldSpec 4 Standard-Res Spectroradiometer. https://www.malvernpanalytical.com/en/products/product-range/asd-range/fieldspec-range/fieldspec-4-standard-res-spectroradiometer. Accessed 15 April 2018.

Mehmood, T., Hovde Liland, K., Snipen, L., & Saebo, S. (2012). A review of variable selection A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 118(0) 62-69.

Mevik, B. H., Wehrens, R., & Liland, K., H. pls: Partial Least Squares and Principal Component Regression. R Package Version 2.6-0, 2016, https://CRAN.R-project.org/package=pls

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., & Mouazen, A. M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil & Tillage Research*, 155, 510-522.

Reeves III, J.B., McCarty, G.W., & Meisinger, J.J. (1999). Near infrared reflectance spectroscopy for the analysis of agricultural soils. *Journal of Near Infrared Spectroscopy*, 7, 179– 193.

Reeves III, J.B., & McCarty, G.W. (2001). Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fibre-optic probe. *Journal of Near Infrared Spectroscopy*, 9(1), 25-34.

Rinnan, A., van den Berg, F., & Engelsen, S.B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28, 1201–1222.

Savitzky, A.; Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem., 36, 1627–1639.

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Witterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, 107, 163–215.

Stevens, A., & Ramirez-Lopez, L. prospectr: Miscellaneous functions for processing and sample selection of vis-NIR diffuse reflectance data. R Package Version 0.1.3, 2014. https://CRAN.R-project.org/package=prospectr

Streukens, S., & Leroi-Werelds, S. (2016). Bootstrapping and PLS-SEM: A step-by-step guide to get more out of your bootstrap results. *European Management Journal*, 34, 618-632.

Veris ® Technologies. P4000, Soil Sensing, Soil Coring. Product Bulletin.

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., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158, 46-54.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J. & Lobsey, C. (2011). Chapter Five – Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. *Advances in Agronomy*, 113, 243-291.

Viscarra Rossel, R.A., Bouma, J. (2016). Soil sensing: A new paradigm for agriculture. *Agricultural Systems*, 148, 71-74

Vohland, M., & Emmerling, C. (2011). Determination of total soil organic C and hot water-extractable C from VIS-NIR soil reflectance with partial least squares regression and spectral feature selection techniques. *European Journal of Soil Science*, 62(4), 598–606.

Vohland, M., Ludwig, M., Thiele-Bruhn, S., & Ludwig, B. (2014). Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, 223-225, 88-96.

Walvoort, D.J.J., & McBratney, A.B. (2001). Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. In Grenier, G., Blackmore, S. (Eds.), ECPA 2001. Proceedings of the Third European Conference on Precision Agriculture, agro Montpellier, Montpellier, France, pp. 503–507.

Wetterlind, J., Stenberg, B., & Söderström, M. (2010). Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. *Geoderma*, 156, 152–160.

Wetterlind, J., Stenberg, B., & Viscarra Rossel, R. A. (2013). Soil analysis using visible and near infrared spectroscopy. In Frans J.M. Maathuis (ed.), *Plant Minaral Nutrients : Methods and Protocols*, Methods in Molecular Biology (953), pp 95 – 106. New York: Springer.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.

Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration method in chemistry solved by the PLS method. In: Ruhe, A., Kagstrom, B. (Eds.), Proc. Conf. Matri Pencils, Lecture Notes in Mathematics. Springer-Verlag, Heidelberg, 286–293.