Vis/NIR spectroscopic measurement of selected soil fertility parameters of Cuban agricultural Cambisols

Ahmed Chacón Iznaga[a,b,*], Miguel Rodríguez Orozco[a], Edith Aguila Alcantara[a], Meilyn Carral Pairol[a], Yanet Eddith Díaz Sicilia[a], Josse de Baerdemaeker[b], Wouter Saeys[b]

[a]*Central University "Marta Abreu" of Las Villas, Faculty of Agricultural and Animal Sciences, Cuba*
[b]*Department of Biosystems, Division of Mechatronics, Biostatistics and Sensors (MeBioS), Catholic University of Leuven, Kasteelpark Arenberg 30, B-3001 Leuven, Belgium*

*Corresponding author. Tel.: +53 42 28 1692; fax: +53 42 28 1608
E-mail address: ahmedci@uclv.edu.cu

Abstract

The conventional methods frequently used in Cuba to determine some fertility parameters important for sugarcane production, such as organic matter (OM), available phosphorus (P) and potassium (K_2O), are difficult, costly, and time-consuming procedures. This study was undertaken to build and validate Vis/NIR calibration models of these parameters at landscape level and within a field, by taking into consideration their correlation coefficients with the OM. The parameters P and K_2O, which are not spectrally active in the Vis/NIR range should be better predicted when are highly correlated with OM. Also, the wavelength intervals to simplify this methodology were selected. Samples were air-dried before scanning using a diode array spectrophotometer covering the wavelength range from 399 to 1697 nm. The regression models were built by using the linear multivariate regression method Partial Least Squares (PLS), and the nonlinear multivariate regression methods Support Vector Machines (SVM) and Locally Weighted Regression (LWR). At landscape level the best correlations between soil spectra and OM ($0.90 \le R^2 \le 0.93$; $0.12 \le RMSEP \le 0.14$) were obtained with LWR, followed by K_2O with LWR ($0.77 \le R^2 \le 0.79$; $3.47 \le RMSEP \le 3.62$), Olsen P ($0.69 \le R^2 \le 0.81$; $0.27 \le RMSEP \le 0.35$) and Oniani P ($0.64 \le R^2 \le 0.65$; $3.31 \le RMSEP \le 3.61$) both with SVM. Also, the nonlinear regression models gave the best results within a field. The higher values for OM (R^2=0.92; RMSEP=0.14) and Olsen P ($0.68 \le R^2 \le 0.83$; $0.27 \le RMSEP \le 0.34$) were observed with SVM, while for K_2O ($0.16 \le R^2 \le 0.63$; $5.13 \le RMSEP \le 5.88$), and Oniani P ($0.70 \le R^2 \le 0.72$; $2.32 \le RMSEP \le 2.52$) were obtained with LWR. The soil fertility parameters studied at landscape level and within a field were best estimated by using nonlinear regression models.

Keywords: Soil fertility parameters; Spectrally active; Visible and Near infrared; Wavelength intervals

1. Introduction

Sugarcane is the major crop of Villa Clara province located in the central part of Cuba. For growth and development of this crop, available potassium (K_2O) and available phosphorous (P) are the most important soil fertility parameters, while organic matter (OM) is needed for good soil physical properties. The conventional methods frequently used to determine these parameters are often too difficult, costly, and time-consuming for practical use.

According to Moore et al. (1993) and Florinsky et al. (2002), conventional soil surveys (sampling and soil fertility analyses) are not sufficient to obtain all the quantitative information about the spatial distribution of soil fertility parameters that would be needed for optimal fertiliser use. Thus, in most developed countries soil sampling and wet chemical analyses are regularly performed to determine the average soil fertility parameters for an agricultural field. These analysis methods require that the moist soil samples collected from the field must be air-dried as soon as possible before sending to a soil testing laboratory. After the whole process of sieving and grinding the soil samples, wet chemical analyses are performed on them. Then, the final results are communicated to the farmers for fertiliser management decision-making.

Visible (Vis) and Near Infrared Reflectance (NIR) spectroscopy has been suggested to be an efficient tool to predict, within a field, soil fertility parameters that can be of significant value when establishing agricultural field trials and in precision farming (Wang et al., 2008). In the context of precision agriculture, Vis/NIR spectroscopy might be an alternative to the conventional analysis methods employed in the Villa Clara province for determining soil fertility parameters (OM, P and K_2O) using a single spectrum per sample.

In order to achieve this purpose, first it is necessary to calibrate and validate measured laboratory Vis/NIR spectra against laboratory determined soil fertility parameters. However, it is often the case that some soil fertility parameters, such as K_2O and P, cannot be directly determined from the Vis/NIR spectra of soil samples as they are not spectrally active in the Vis/NIR range (Ben-Dor et al., 1997; Malley et al., 2002; Saeys and Ramon, 2004). Therefore, it could be expected that the prediction of these parameters is at least partially based on their correlation with others such as OM content.

Organic matter content is highly correlated with soil spectra, as it is spectrally active in the Vis/NIR range. Therefore, if OM is highly correlated with P and K_2O, then the soil spectral reflectance could be a useful tool for providing a referenced measure for their successful prediction. In addition, by understanding how reflectance in specific wavelength intervals could be related to the soil fertility parameters, it could be used to develop a simple low-cost methodology without using the full spectrum. Thus, Vis/NIR spectroscopy could be a useful tool for on-site analyses of soil samples for soil fertility mapping as input to variable rate fertilisation. If successful this research could promote precision fertilisation in the Villa Clara province, where it has been demonstrated that there is a gap at present between the applied fertiliser rates based on present recommendations and the real requirements of the crop, which implies a significant loss in sugarcane production (De León et al., 2004). Therefore, a tool which allows the assessment of the variation in soil fertility within a field would increase the efficiency of the applied fertiliser. Consequently, it would contribute to reducing the production costs and the negative impacts on the environment related to nutrient losses from the field. By replacing the wet chemical analyses by Vis/NIR spectroscopic scanning in the laboratory, it could be possible to speed up the analyses and thus increase the number of samples that could be be analysed.

The overall goal of the present study was to assess the profitability of Vis/NIR spectroscopy in laboratory conditions, for measuring the average fertility on air-dried soil samples at landscape level and within a field in the Villa Clara province. The correlation coefficients of P and K_2O with the OM were taken into account to achieve this purpose. Also, the wavelength intervals that contribute most to the development of the Vis/NIR regression models were determined for each soil fertility parameter.

2. Materials and methods

2.1. Site description

The Cambisol fields studied are all located in the Villa Clara province in the central part of Cuba between the coordinates 22°16’, 23°09’ N and 80°02’, 80°25’ W. As this study was to test the potential of Vis/NIR spectroscopy for measuring different soil fertility parameters, the main criterion taken into account when selecting this landscape area was based on having a representative calibration set covering the variation in Cambisols from the Villa Clara province. Cambisols are soils at an early stage of soil formation. There is generally a brownish discoloration below the surface horizon to mark the beginning of pedogenesis. The horizon differentiation is weak. Cambisols are developed in medium and fine-textured materials derived from a wide range of rocks, mostly in alluvial, colluvial and aeolian deposits (FAO, 2001).

The productive capacity of the agricultural soils used for sugarcane production in the Villa Clara province can be directly related to the improvement of soil fertility parameters such as OM, P and K_2O. In this regard, satisfactory crop production is obtained only when the soil possesses favourable soil fertility parameters. Nutrients can be directly added by the application of mineral fertiliser to the soil. However, the addition of fertiliser alone is not enough to retain a sufficient level of soil fertility (van Schöll and Nieuwenhuis, 2004). In this sense, most fertilisers commonly used in agriculture contain the three basic plant nutrients (N, P, K), for instance, urea (46% N), triple superphosphate (46% P_2O_5) and potassium chloride (60% K_2O). However, the management based on the average soil fertility is no longer enough to maintain sustainable high yields in the agricultural production.

Fertiliser can be applied in a number of ways such as broadcasting (the fertiliser pellets are spread evenly over the whole field, and then often ploughed or raked into the soil) and row application (the fertiliser is applied in rows). In this way, several authors such as Olsen *et al*. (1971), Rowse and Stone (1980), and Mandal and Thakur (2010) reported that only 40 to 50% of N fertilisers and 20 to 30% of P and K fertilisers are effectively used by crops. Thus, the remaining % become volatilised, leached to groundwater, or get fixed within the soil. Also, broadcasting of fertilisers, especially P and K, produces fixation problems due to greater soil contact, whereas volatilisation of N results in a decrease in the amount of N available content in the soil. Applying fertiliser in this manner results in lower productivity and profitability due to missing out on additional yield in the parts of the field that are under-fertilised and further reduced profitability where fertiliser is over-applied (Phillips, 2009).

2.2. Experimental fields and procedure

A total of 189 soil samples were collected on 189 fields ranged from 2.2 to 2.8 ha. These fields are located on 10 Agroindustrial Complexes (CAI, Spanish acronym). Cambisols are present at all sampling locations and show a different proportional distribution. The variability in soil fertility parameters present in these areas on Cambisol soils can be a factor in the spatial variability in sugarcane yield (Table 1) as Pérez et al. (2001) observed.

The soil samples were taken from the plough layer (0 – 20 cm) following the sampling scheme proposed by the Fertilizers and Amendments Recommendations Service (SERFE, Spanish acronym). In this system each soil sample consisted of 30 subsamples which were taken from subplots located across a diagonal line on the field, which starts and finishes 10 m from the field borders. The soil sampling was done before planting when the fields were ready to be furrowed. The samples were collected using a sampling auger with a footrest.

Also, a total of 37 field subsamples were selected from one field of 2.03 ha. This field is within an experimental area of the Central University "Marta Abreu" of Las Villas. These samples were used for establishing an independent validation set in order to assess the accuracy of the Vis/NIR calibration built based on samples from other fields, for evaluating the soil fertility variation within an individual field. The soil samples were collected at the same depth (0 – 20 cm) by using a random sampling scheme.

Each sample was bagged separately into plastic bags with identification labels, air-dried at room temperature to constant weight, sieved with a 0.5 mm sieve and homogenised before Vis/NIR spectroscopic measurement and chemical analysis.

2.3. Chemical analyses and procedure

The collected soil samples were analysed for OM, K_2O and P. These soil fertility parameters are some of the most important for sugarcane growth. Knowledge of the spatial variability of these soil fertility parameters and analyses of the spatial correlation between them are important for sugarcane production. Recognising the significance of quantifying and managing this variability that occurs in agricultural fields could lead to different approaches for implementing a site-specific management in sugarcane production in Villa Clara province. Also, these are the most commonly analysed parameters within the SERFE system.

The chemical analyses were done in the analytical chemistry laboratories of the Territorial Station for Sugar Cane Research (ETICA, Spanish acronym) and in the Agricultural Research Center (CIAP, Spanish acronym) belonging to the Central University "Marta Abreu" of Las Villas (applying the same methods used by SERFE). The methods used for conventional chemical analyses are summarised in Table 2.

Basic descriptive statistics (range, mean, median, skewness, kurtosis, minimum, maximum, standard deviation, coefficient of variation) and Pearson correlations were obtained by processing data with the statistical functions included in the Analyses Toolpack of Microsoft Excel 2007. Standard error of laboratory (SEL) was calculated as:

$$SEL = \sqrt{\frac{\sum_{i=1}^{N}[\sum_{j=1}^{R}(y_{ij}-\bar{y}_j)^2/(R-1)]}{N}} \qquad (1)$$

where y_{ij} is the *j*th replicate of the *i*th sample, $\bar{y}_j$ is the reference method mean value of all the replicates of the *i*th sample, *N* is the number of samples, and *R* is the number of replicates.

In order to compare the variability of the soil fertility parameters among themselves across the landscape and within a field, the coefficient of variation (CV) was used. The results were categorised into the three classes proposed by Aweto (1982), where CV < 25% = low variability, 25 < CV < 50% = moderate variability, 50 < CV <100% = high variability.

2.4. Spectral data acquisition

The setup for acquiring Vis/NIR diffuse reflectance spectra of the air-dried soil samples in the laboratory consisted of a diode array spectrophotometer (CORONA PLUS REMOTE Vis/NIR SB, Zeiss, Jena, Germany) and an OMK 500-H measuring head connected to it with an optical fibre-bundle. The Aspect –*plus* software supplied by Carl Zeiss Jena GmbH was used to control the spectrophotometer and collect soil spectra.

All spectra were obtained using the same instrument settings: average 10 scans per spectrum measured in reflection mode, for wavelength range from 399 to 1697 nm. The integration times for the Vis (Si: 400 – 1000 nm) and the NIR (InGaAs: 900 – 1700 nm) ranges were set to 54 ms and 43 ms, respectively. The spectral resolution in Vis was 3.3 nm and in NIR was 10 nm.

A small amount of the soil sample (about 30 g) was placed in a petri dish of 10 mm depth and 35 mm diameter. The soil in the petri dish was first compacted and then carefully levelled in order to obtain a smooth surface to reduce variation due to the packing of the soil. Each petri-dish with 30 g of soil was placed under the soil sensor at the focal point. Then three reflectance spectra were taken over the central area of the petri dish rotating the sample approximately 120° between each spectral acquisition. The three spectra of each soil sample were averaged to obtain one average spectrum per sample. Each spectrum was saved as an individual file (.csv), and then these files were assembled into a single matrix to be imported into MATLAB.

2.5. Prediction of soil fertility parameters from the measured spectra

Prediction models linking the acquired Vis/NIR reflectance spectra to the selected soil fertility parameters were built using the linear multivariate regression method Partial Least Squares (PLS), and the nonlinear multivariate regression methods Support Vector Machines (SVM) and Locally Weighted Regression (LWR). The PLS regression relates the variations in one response variable to the variations of several predictors (wavelengths) as stated by Yang et al. (2011). This method is based on orthogonal transformation technique. It reduces the complexity of modelling and eliminates the adverse effects of multicollinearity among spectral variables. SVM focuses on minimising a bound on the risk function, rather than minimising the error in training data. In this way, the over-fitting problem is prevented (Karimi et al., 2008). For nonlinear cases, SVM uses a so-called kernel technique to plot the data into a higher dimensional feature space, where linear functions can be applied. More detailed information on SVM may be found in Vapnik (1995), Vapnik et al. (1997), Smola and Scholkopf (1998). In LWR, the spectra is first compressed by using principal component analysis, and then the Mahalanobis distance is computed on the first principal components which accounts for a given percentage of cumulative explained variance. After this procedure local PLS calibrations for each unknown sample are carried out in the spectral space using its k-nearest neighbours which are weighted according to their distance from the unknown sample (Ramirez-Lopez et al., 2013).

For the Vis/NIR prediction at landscape level, the 189 samples taken from the different Cambisol fields in the Villa Clara province were randomly divided into a calibration set of 126 samples and a validation set of 63 samples. For the prediction within a field, the 189 soil samples from different agricultural fields were used as a calibration set. The test set consisted of 37 Cambisol soil samples sampled separately from one field which had not been included in the set of 189 calibration samples.

During the analyses, several samples were identified as outliers, or strange values whose presence could alter the results in a remarkable way. The criterion for identifying an outlier was based on the examination of the data for unusual observations which do not conform to the pattern established by other observations. Outlier analysis was also performed by screening from the soil data sets those data for which measured concentrations of soil fertility parameters were significantly higher or lower than the sample population and therefore could have been influenced by the sampling and chemical analysis procedures.

All calculations were performed in MATLAB 7.9 (R2009b, The Mathworks, Nattick, MT). The Venetian blinds cross-validation strategy (10 splits) was used to evaluate the prediction error as a function of the model complexity (Leung, 2005).

In order to select the most suitable pre-processing methods for soil spectra, a trial and error process was followed. Depending on the combination of preprocessing methods, differences regarding to several statistics were observed. Therefore, the selection of the best pre-processing method was related to the prediction performance of the calibration models. This performance was evaluated based on the calibration statistics: coefficient of determination R^2 in cross-validation and test set prediction, root mean square error of cross validation RMSECV and test set prediction RMSEP. Also, the statistics R^2 in test set prediction, RPD (ratio of performance deviation) and RER (ratio of error range) were used for comparing the calibration accuracies obtained for the different soil fertility parameters. The RPD is a statistic calculated by:

$$RPD = \frac{SD}{SEP} \qquad (2)$$

The RER is given by:

$$RER = \frac{Range}{SEP} \qquad (3)$$

Calibration accuracy was assessed using the guidelines proposed by Malley et al. (2004) and Nduwamungu et al. (2009) for environmental samples like soil. These criteria are summarised in Table 3.

After the Vis/NIR prediction within a field, a kriging interpolation algorithm implemented in the software MATLAB 7.9 (R2009b, The Mathworks) was used for mapping these soil fertility parameters. Maps were comparing by means of the coefficient of variation (CV) of measured and predicted values.

2.6. Wavelength selection

Data were analysed with Forward Interval PLS model (iPLS) to determine those specific wavelength regions important for better estimating each soil fertility parameter. This method is based on the division of the full spectrum into smaller and equidistant intervals (15) of equal width (28 variables). The RMSECV is calculated for each interval and compared with the value obtained for the full spectrum model. Regions that present the smallest value of RMSECV are then chosen (Müller et al., 2011; Nørgaard et al., 2000). The calculations were performed in MATLAB 7.9.

3. Results and discussion

3.1. Spatial variability of soil fertility parameters

The descriptive statistics of the soil fertility parameters for the Cambisol soil samples from different fields in the Villa Clara province are listed in table 4. The higher coefficients of variation (CV) were observed in K_2O (36.49%) and Oniani P (44.62%). From the CV limits used in this research both parameters had a moderate variability (25 < CV < 50%) across the landscape. On the other hand the lowest CV was observed for OM (14.08%). This extensive degree of spatial variability provides the opportunity to apply site-specific fertiliser management strategies to reduce misapplications by improving the match between plant fertiliser requirements and fertiliser supply.

Table 5 shows the results of soil fertility parameters analysed within a Cambisol field. According to the CV limits used in this research, none of these parameters had a high variability (50 < CV <100%). However, K_2O and Oniani P showed a moderate variability with 37.68%

and 31.70% respectively. On the other hand the CV values for OM and Olsen P were less than 16%, which indicated a low variability for these parameters.

3.2. Correlation between soil fertility parameters

Consistent and positive correlations were found between all the considered soil fertility parameters at landscape level. Another significant aspect was that the higher correlation coefficients observed for K_2O, Olsen P and Oniani P were related to OM (Table 6), with the highest correlation coefficients between OM and phosphorus, at 0.89 for Olsen P and 0.88 for Oniani P. The lowest correlation was between Olsen P and K_2O at 0.68, which is still significant.

The results of the correlation analyses presented in Table 7 indicate that all the measured soil fertility parameters within a field exhibited positive correlations of diverse magnitude. These correlations were very similar to those obtained for the Cambisol soil samples at landscape level. Similar to the landscape level, the highest correlation coefficients observed for K_2O, Olsen P and Oniani P were related to OM. According to Hodges (2010) higher OM levels can help reduce P fixation reactions, by binding Al, Fe and Ca, and forming soluble complexes with P. Another similarity to the landscape correlations was that the highest correlation coefficient was observed between OM and Olsen P at 0.85. Again, the lowest value was observed between Olsen P and K_2O at 0.65.

3.3. Prediction of soil fertility parameters at landscape level

The final selected pre-processing method included logarithm transformation [Log (1/R)], smoothing and dataset centring (Mean Centre). Spectral preprocessing with mathematical functions is commonly used to correct for non-linearities and electronic noise of the detector. Also, pre-processing methods aim to remove variation in the spectra which is not caused by the component of interest, but due to light scattering and chemical interference.

In Fig. 1 the diffuse reflectance spectra of Cambisol at landscape scale (calibration and validation set) before and after pre-processing are shown. These curves (left) are a graphical representation of the spectral reflectance of the air-dried soil samples used in this research, as a function of the studied wavelength range in the Vis/NIR region. In the visible region (380 – 780 nm) the reflectance is lower than in the NIR region (780 – 1697 nm). Also, around the wavelength of 1400 nm, the reflectance values show a local minimum. The absorption of light observed in the 1350 to 1450 nm region is primarily by O-H, C-H and N-H bonds. Stenberg et al. (2010) explained that the absorptions in the NIR region result from the overtones of OH, SO_4, and CO_3 groups, as well as combinations of fundamental features of H_2O and CO_2.

The performance of the selected models at landscape level is illustrated in Fig. 2, which shows the correlations between the measured values and those predicted by the Vis/NIR model. In cross validation, the best prediction coefficient (R^2) was obtained for OM (0.93) with LWR regression analysis. Also, for K_2O this coefficient (0.79) was better with the same nonlinear regression model (LWR).

For OM the data points were closely grouped around the target line both in the calibration set and in the prediction set, while they were slightly more scattered around this line in both sets for Olsen P and Oniani P; for K_2O a higher scattering was observed. By comparing the SVM prediction models for Olsen P and Oniani P, it was observed that the prediction performance for Oniani P was considerably worse (R^2 of 0.81 vs. 0.65). This suggests that the P content measured by Vis/NIR spectroscopy corresponds better to the P content assessed with the Olsen method than with the Oniani method. A possible explanation for this may be found in the fact that the Olsen method for determining the available P content shows a better performance in calcareous and neutral soils such as Cambisols. The Oniani method, when used in calcareous soils for analysing the available P content, tends to dissolve calcium carbonate ($CaCO_3$).

Table 8 shows the results of the better regression models and others used for the prediction of the soil fertility parameters at landscape level. The three regression models (PLS, SVM, LWR) developed for OM achieve a successful prediction performance. However, LWR produces more accurate results with the highest R^2 of prediction (0.93) and the lowest RMSEP (0.12). Also, LWR model shows a minimum difference (0.01) between RMSEC and RMSECV; however there was not difference between these statistics for the other two regression models. The higher accuracy obtained for OM was based on the characteristic of LWR in selecting the most suitable soil samples for this soil fertility parameter. In line with Stenberg et al. (2010), OM is one of the fundamental constituents of the soil which has well-recognisable absorption features in the Vis/NIR region. This indicates that the prediction of this constituent in new Cambisol samples based on the Vis/NIR spectra will be reliable.

Similar statistical results for K_2O were obtained from PLS, LWR and SVM. In all cases the prediction accuracy is classified as moderately useful ($0.70 \leq R^2 < 0.80$). This soil fertility parameter is not spectrally active in the Vis/NIR range. Therefore it was estimated indirectly due to its high correlation (r=0.78) to a more spectrally active parameter such as OM, which influences most soil chemical properties. The soil OM can be directly related to the absorption in Vis/NIR spectra through a number of functional groups such as NH, CH, and CO groups. In this sense Irons et al. (1989) considered that an increase in the OM content of a soil generally causes a decrease of reflectance over the entire spectrum. A high OM content and hence, a strong decrease of overall reflectance, might even mask other absorption features in the soil spectra.

Olsen P was better predicted with SVM (R^2= 0.81). As the R^2 value in prediction was equal to that obtained in cross-validation, this model is expected to be quite robust. For this fertility parameter the prediction performance of the PLS regression model (R^2 of 0.69) was considerably worse than that for the other two regression models (LWR and SVM, both with R^2 of 0.81). This suggests that the relation between the preprocessed absorbance spectra and the Olsen P is nonlinear. Thus, SVM and LWR models are more reliable for the prediction of nonlinear data with a moderately successful accuracy for this fertility parameter. Also, according to Borggaard (2004), a potential drawback with using PLS regression and all linear regression techniques is the possible nonlinearity of correlations between NIR spectra and the property of interest. This fertility parameter, just like K_2O, is not spectrally active in the Vis/NIR range. Therefore, the prediction accuracy for Olsen P could be substantially improved because this parameter is also highly correlated (r=0.89) with OM. In the case of Oniani P, all the models showed similar R^2 to each other ($0.64 \leq R^2 \leq 0.65$). According to the criteria suggested by Malley et al. (2004), this prediction accuracy is classified as less reliable. The prediction accuracy for P is considerably lower for Oniani than for Olsen method. The differences between both analytical methods resulted in different proportions of this soil fertility parameter.

In general for all soil fertility parameters, equally good or better prediction performance was obtained with the nonlinear regression models (LWR and SVM) than with the linear regression model (PLS). However, as already mentioned for OM also successful prediction results were obtained with PLS. This parameter showed a nonlinear relation with the others. Janik and Skjemstad (1995) suggested that nonlinear relationships between spectra and soil variables often occur with data covering wide ranges and can be the result of distortions of strong signals, and different mineralogical values with high and low soil variable values in the calibration data sets.

These results were in line with those obtained by Shao and He (2011). These authors reported R^2 equal to 0.82 and 0.80 for phosphorus and potassium respectively. In that case they used the Bray and Kurtz method for analysing phosphorus and flame atomic emission spectrometry for potassium. As the K and P content were found to be highly correlated to the OM content, the prediction of the P and K content in soil from Vis/NIR spectra may be based on their correlation with the OM content.

3.4. *Prediction of soil fertility parameters within a field*

The raw and preprocessed data for the different Cambisol samples taken from one field are illustrated in Fig. 3. In line with the results obtained at landscape level the final selected pre-processing method was Log (1/R), Smoothing and Mean Centre.

The reflectance spectra at landscape level (a) and within a field (b) have the same general characteristics, which suggest that the soil type could be derived from the spectra.

In Fig. 4 the scatter plots of the better predicted vs. measured values of all the soil fertility parameters within a field are shown. Across the full range, OM content (SVM, R^2 of 0.92) exhibited a good agreement between calibration and validation data at lower and medium values than at higher values of the validation set. On the other hand for K_2O (LWR, R^2 of 0.63) the data points tend to be scattered along the whole range of the model. The best regression models obtained for Olsen P (SVM, R^2 of 0.83) and Oniani P (LWR, R^2 of 0.72) illustrated more accurate predictions for these parameters. These higher predictions could also be explained by the relatively higher correlations of these soil fertility parameters with OM (r = 0.85 and 0.83 for Olsen P and Oniani P, respectively).

By comparison, the nonlinear regression models produce better predictions within a field as well as at landscape level. It is worth noting that SVM model for OM achieves similar accuracy within a field (R^2 of 0.92; RMSEP of 0.14) than LWR at landscape level (R^2 of 0.93; RMSEP of 0.12). LWR model for K_2O shows a lower prediction within a field (R^2 of 0.63; RMSEP of 5.13) than at landscape level (R^2=0.79; RMSEP=3.47). Also, the prediction of the best models (SVM) for Olsen P produces moderately successful performance in both scenarios (R^2 of 0.83; RMSEP of 0.27 within a field and R^2 of 0.83; RMSEP of 0.27 at landscape level). Finally, the numerical values of statistics for Oniani P obtained with SVM model within a field (R^2 of 0.72; RMSEP of 2.32) are better than those obtained with LWR model at landscape level (R^2 of 0.65; RMSEP of 3.31). Further, when the best models are compared the results demonstrate that for this soil type a calibration made within a field is capable of being used over the landscape scale.

The statistics on the prediction performance of Vis/NIR spectroscopy using PLS, LWR and SVM for the Cambisol field are summarised in Table 9. The best test set prediction results were obtained for organic matter (R^2 of 0.92), while the results for K_2O were worst ($0.61 \leq R^2 \leq 0.63$). The obtained prediction R^2 were $0.68 \leq R^2 \leq 0.83$ for Olsen P and $0.70 \leq R^2 \leq 0.72$ for Oniani P.

In general LWR was found to give the best prediction results. In line with Volkan et al. (2010) the values of R^2 for soil fertility parameters prediction in the validation set were lower and RMSEP values higher than corresponding values in the calibration set. Except for Olsen P the R^2 values were slightly worse in test set prediction than in cross-validation. This could have been expected as the variation within a field is smaller than the variation between different fields at landscape level and some of the unspecific correlations which exist between the spectra and chemical constituents of different fields may not exist within a field (Dardenne et al., 2000; Volkan et al., 2010).

According to the suggested guidelines, there was a coincidence in the predictive accuracy of the best regression models for OM and Olsen P; both were classified as moderately successful. For K_2O the model was less reliable, which could be explained by the lack spectral activity of this parameter.

Finally, the prediction of Oniani P was classified as moderately useful. Bogrekci and Lee (2007) also reported good prediction potential of the P concentration from Vis/NIR reflectance spectra, with R^2 values of 0.93, 0.95 and 0.76 for total, Mehlich-1 and water-soluble P, respectively. In this research the very well predicted P content (Olsen and Oniani) by using Vis/NIR spectroscopy might be related to the higher correlation with OM.

On the other hand, the results for OM and K in this research were very similar to those obtained by He et al. (2007). They achieved good predictions with PLS for OM (R^2 of 0.93). However, they concluded that Vis/NIR spectroscopy was not a good tool for P and K prediction with R^2 values of 0.47 and 0.68, respectively. Also, in the study carried out by Wetterlind et al. (2008) the validation statistics indicated that the Vis/NIR calibrations for OM were reliable (R^2 of 0.89 and R^2 of 0.87).

In Fig. 5 and Fig. 6 the interpolated maps of the wet chemically measured and Vis/NIR predicted values of the soil fertility parameters for the Cambisol field are presented. Moreover, as the prediction performance is less accurate, the CV between the wet chemically measured and Vis/NIR predicted maps increase.

For OM (R^2= 0.92 – successful accuracy) the maps are fairly similar, according to the distribution of different colours, which indicates that the distribution of the soil fertility content has been quite well captured by the OM values predicted based on the Vis/NIR spectra. From the CV limits used in this research OM showed a lower variability in both maps, with a similar CV of 12.16% and 12.34% for measured and Vis/NIR predicted values respectively.

In case of K_2O (R^2= 0.63 – less reliable accuracy) a remarkable difference can be observed between the measured (CV of 37.68%) and Vis/NIR predicted (CV of 19.88%) values.

The maps obtained for Olsen P (R^2= 0.83 – moderately successful accuracy) are also comparable. In this case, the CV values (15.27% for measured and 12.34% for Vis/NIR predicted values) indicate a lower variability.

Finally, the measured and Vis/NIR predicted values for Oniani P (R^2= 0.72 – moderately useful) showed a moderate variability, but the differences between them were higher than the observed for OM and Olsen P. This variability exhibited a CV of 31.70% and 26.70% for the measured and Vis/NIR predicted values respectively.

3.5. Identification of effective wavelength intervals for each soil fertility parameter

The results of the above regression models were based on the full spectrum, which included a large number of wavelengths. Therefore, the most practical approach to simplify this methodology is to select the wavelength intervals with the lowest RMSECV that contribute most to the development of the Vis/NIR regression models. The number of wavelengths required was reduced considerably to a maximum of 84 (each interval includes 28 wavelengths), which represent 20% of the full spectrum. Thus, different wavelength intervals were selected as important for each soil fertility parameter at landscape level, but some intervals were common in some of these parameters (Fig. 7).

The interval number 1 (399 – 489 nm) was common to all the soil fertility parameters at landscape level. In addition, the interval number 2 (489 – 580 nm) and the interval number 3 (580 – 674 nm) were selected for K_2O and Olsen P. All these intervals fit in the Vis region (380 – 780 nm). In general, the absorptions in this region are mainly related to those minerals which contain iron. Also, in other cases soil OM has a tendency to show broad absorption peaks in the Vis region, which are dominated by chromophores (spectrally active groups e.g. Fe, OH^- in water and minerals, CO_3^{2-}, Al^{2+}, SO_4^{2-} in minerals). This statement is in agreement with Clark et al. (1990) and Clark (1999). The other two intervals selected for OM (1320 – 1402 nm; 1480 – 1555 nm) are included in the NIR region (780 – 2500 nm).

Wavelengths included in the intervals number 1 and 11 selected for OM are in line with the results reported by several authors. For instance, Lee et al. (2003) estimated chemical properties in Florida and found that the correlation coefficients between the OM content and spectral reflectance in 428 nm and 1376 nm were higher. Mapping soil OM in the north-west part of Semnan province (Iran), Nowkandeh et al. (2013) selected 477 nm, as one of the best wavelengths for regression modelling. In Shan Dong province (China), Wang et al. (2013) reported that two of the optimal wavelengths which have the best fitness for predicting OM were 399 nm and 449 nm. Also, Yu et al. (2013) reported for OM content that the wavelength of 492 nm and 1317 nm possessed the best prediction accuracy. Both wavelengths are nearest of 489 nm (interval number 1) and 1320 nm (interval number 2) respectively.

On the other hand, wavelengths included in the intervals selected in this study for K_2O, Olsen P and Oniani P have been reported by other researchers. In research carried out in two agricultural fields in north-eastern Mississippi, Thomasson et al. (2001) selected the wavelengths 425 nm, 525 nm, 575 nm, 625 nm and 675 nm for potassium (K). The last wavelength is close to 674 nm, which is the boundary of the interval number 3 of this fertility parameter. Also, for phosphorus (P) these authors selected the wavelengths 475 nm, 525 nm, 575 nm, 625 nm, 1075 nm, 1125 nm, 1225 nm, 1425 nm, 1475 nm and 1525 nm. For K, Lee et al. (2003) reported the wavelengths 428 nm, 444 nm and 522 nm. While for P the same authors reported the wavelengths 428 nm, 430 nm, 522 nm, 602 nm, 612 nm and 1100 nm.

Table 10 shows the results of the predictive regression models at landscape level based on the selected wavelength intervals. The three regression models (PLS, SVM, LWR) developed for OM achieve an excellent prediction performance. It was a better result than the successful accuracy obtained with the full spectrum. Also, LWR model gave the more accurate results with the highest R^2 of prediction (0.97) and the lowest RMSEP (0.08). The other soil fertility parameters also increased the R^2 of prediction and kept the model selected as the best with the full spectrum (LWR – K_2O; SVM – Olsen P and Oniani P). However, only K_2O and Oniani P differ as to the category related to the accuracy of performance observed with the full spectrum. The new accuracy of performance for K_2O achieves R^2 of 0.82 (moderately successful) and for Oniani P R^2 of 0.70 (moderately useful). The results demonstrate that all models developed for the soil fertility parameters using just 20% of the wavelengths performed better than with the full spectrum. Therefore, the highest accuracy levels and the lowest RMSEP obtained with a small number of wavelengths simplify and make more practical the methodology when applied at landscape level.

Fig. 8 demonstrates that within a field, as well as at landscape level, the wavelengths included in interval number 1 were selected for all fertility parameters. Also, there were coincidences as to the three intervals selected for K_2O (number 1; 2; 3) and in two of the intervals selected for Oniani P (number 1; 8). Therefore, some wavelengths selected at landscape level and within a field fall in comparable spectral regions.

Nowkandeh et al. (2013) found 477 nm, 905 nm, 972 nm, 1013 nm, 1023 nm and 1033 nm as most important wavelengths for detecting OM. The first wavelength is included in the interval number 1 and it is in agreement with those observed within a field. The second one belongs to the interval number 6 (860 – 952 nm) and the rest to the interval number 7 (952 – 1050 nm). For the other soil fertility parameters, some of the wavelengths selected by Thomasson et al. (2001) and Lee et al. (2003) are in agreement in both scenarios too. For K_2O there is a total coincidence at landscape level and within a field. For P, there was a coincidence with the wavelengths 428 nm, 430 nm, 475 nm, 1075 nm, 1100 nm and 1125 nm. However, these authors also reported other wavelengths for P (1225 nm, 1412 nm, 1425 nm, 1475 nm, 1498 nm, 1525 nm) which are in agreement with those observed in this study within a field but differ from the wavelengths selected at landscape level. Moreover, the regions included in the intervals selected, are known to correspond to spectral features related to these soil fertility parameters.

In table 11 the results of the predictive regression models within a field based on the selected wavelength intervals are shown. The three regression models (PLS, SVM, LWR) developed for OM achieve a successful prediction performance. It was the same result obtained with the full spectrum within a field. However, the RMSEP values were increased as a consequence of reducing the number of wavelengths. Both results differ from those obtained at landscape level, where the three regression models improve the prediction accuracy and reduce the RMSEP values.

By reducing the number of wavelengths, the soil fertility parameters K_2O and Olsen P increased the R^2 of prediction, reduced the RMSEP and kept the model selected as the best with the full spectrum (LWR – K_2O; SVM – Olsen P). Also, the prediction accuracy was the same as observed with the full spectrum, except for LWR (R^2 of 0.82 – moderately successful). For Oniani P the R^2 of prediction and RMSEP values were increased. The best model was SVM and not LWR as obtained with the full spectrum.

The RMSEP values were increased only within a field and just for OM and Oniani P. In this sense Fig. 8 shows that the wavelength intervals identified were not identical for OM (interval number 6) and Oniani P (interval number 8). It means that a higher RMSECV value than in the others was obtained in this respective interval. It is probable that using data only from the two intervals with a lower RMSECV would give better results of RMSEP for both soil fertility parameters. This verification would be an important additional step toward the improvement of the methodology which pursues the use of a low number of wavelengths for obtaining better or similar prediction accuracy.

4. Conclusions

The potential of Vis/NIR spectroscopy for prediction the average soil fertility parameters of Cuban agricultural fields with Cambisol soil has been evaluated. One preprocessing method has been applied for this purpose.

The soil fertility parameters studied at landscape level were estimated by using nonlinear regression models with a successful (OM; R^2 of 0.93 with LWR), moderately successful (Olsen P; R^2 of 0.81 with SVM), moderately useful (K_2O; R^2 of 0.79 with LWR) and less reliable (Oniani P; R^2 of 0.65 with SVM) accuracy for the Cambisol soil samples. The soil fertility parameters K_2O, Olsen P and Oniani P are not spectrally active in the Vis/NIR range, and then they were most likely predicted through their strong correlation with the OM. The OM content can be predicted from the Vis/NIR spectra thanks to the spectral activity of the CH-bonds.

In general, results within a field indicated that all the regression models (PLS, LWR and SVM) provided good correlations between soil spectra and OM. The SVM regression model gave the best results for OM (R^2 of 0.92) and Olsen P (R^2 of 0.83). The best correlation between soil spectra and Oniani P was obtained with LWR (R^2 of 0.72). These better predictions obtained for phosphorus content (Olsen and Oniani) could be related to their higher correlations with OM. The lower prediction R^2 values for K_2O (R^2 of 0.63 with LWR) might be explained the lack of spectral activity of this molecule.

This paper has also shown that the PLS method consistently identified spectral regions of interest for better estimating each soil fertility parameter. The advantage of the selection of the wavelength intervals is that the full spectrum need not be used. Thus, the number of wavelengths was reduced to 84 which is the 20% of the full spectrum. In general, different wavelengths were found to be important for the different soil fertility parameters. There were some wavelengths in common among K_2O and Olsen P (399 – 674 nm) at landscape level, among all soil fertility parameters (399 – 489 nm) at landscape level and within a field.

Data from other intervals would provide similar levels of accuracy. Therefore, verification would be an important additional step toward to improve this methodology. Further study will focus on the improvement of the Vis/NIR spectroscopy methodology considering different soil types.

References

Aweto A O (1982). Variability *of* upper slope soils development under sandstones in south-western Nigeria. Geographical journal, 25, 27-37.

Ben-Dor E; Inbar Y; Chen Y (1997). The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. Remote Sensing of Environment, 61, 1-15.

Bogrekci I; Lee W S (2007). Comparison of ultraviolet, visible, and near infrared sensing for soil phosphorus. Biosystems Engineering, 2 (96), 293-299.

Borggaard, C. (2004). Neural networks in near-infrared spectroscopy. In: Williams, P., et al. (Eds.) *Near-Infrared Technology in the Agricultural and Food Industries.* Minnesota: American Association of Cereal Chemists, Inc. pp. 101-107.

Clark R N (1999). Spectroscopy of rocks and minerals and principles of spectroscopy, in: Remote Sensing for the Earth Sciences (A N Rencz, Ed.), 3-58. John Wiley & Sons, Chichester, UK.

Clark R N; King T V V; Klejwa M; Swayze G A; Vergo N (1990). High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research, 95, 12653–12680.

Dardenne P; Sinnaeve G; Baeten V (2000). Multivariate calibration and chemometrics for near infrared spectroscopy: Which method? Journal of Near Infrared Spectroscopy, 8, 229-237.

De León M E; Campos J; Betancourt G; Machado I; Soler L (2004). Control lote generalization from the Fertilizers and Amendments Recommendations Service, as a way to increase the sugarcane yield. In: XV FORUM of Science and Technique, Havana City, July 8, 28.

FAO (2001). Lecture notes on the major soils of the world. World Soil Resources Reports, Rome. p. 334.

Florinsky I V; Eilers R G; Manning G R; Fuller L G (2002). Prediction of soil properties by digital terrain modelling. Environmental Modelling & Software, 17, 295-311.

He Y; Huang M; Garcia A; Hernández A; Song H (2007). Prediction of soil macronutrients content using near-infrared spectroscopy. Computers and electronics in agriculture, 2 (58), 144-153.

Hodges S C (2010). Soil Fertility Basics, Soil Science Extension, North Carolina State University.

Janik L J; Skjemstad J O (1995). Characterization and analysis of soils using mid infrared partial least-squares. 2. Correlations with some laboratory data. Australian Journal of Soil Research, 334, 637-650.

Karimi Y; Prasher S O; Madani A; Kim S (2008). Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. Canadian Biosystems Engineering, 50, 713-720.

Lee W. S., Sánchez J. F., Mylavarapu Y.R.S., Choe J. S. (2003). Estimating chemical properties of Florida soils using spectral reflectance. Transactions of the ASAE, 46 (5), 1443-1453.

Leung D H Y (2005). Cross-validation in nonparametric regression with outliers. The Annals of Statistics, 33 (5), 2291-2310.

Malley D F; Yesmin L; Eilers R G (2002). Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy. Soil Science Society of America Journal, 66, 1677-1686.

Malley D F; Martin P D; Ben–Dor E (2004). Application in analysis of soils, in: Roberts C A; Workman J; Reeves J B (Eds.). Near-Infrared Spectroscopy in Agriculture. ASA, CSSA, and SSSA, Madison, WI. pp. 729-784.

Mandal S; Thakur T C (2010). Design and development of subsoiler-cum-differential rate fertilizer applicator. Agricultural Engineering International: CIGR Journal, 12(1), 74 - 83.

Moore I D; Gessler P E; Nielsen G A; Peterson G A (1993). Soil attribute prediction using terrain analysis. Soil Science Society of America Journal, 57, 443-452.

Müller A L H; Flores É M M; Müller E I; Silva F E B; Ferrão M F (2011). Attenuated Total Reflectance with Fourier Transform Infrared Spectroscopy (ATR/FTIR) and Different PLS Algorithms for Simultaneous Determination of Clavulanic Acid and Amoxicillin in Powder Pharmaceutical Formulation. Journal of the Brazilian Chemical Society, 22 (10), 1903-1912.

Nduwamungu C; Ziadi N; Tremblay G F; Parent L E (2009). Near-infrared reflectance spectroscopy prediction of soil properties: Effects of sample cups and preparation. Soil Science Society of America Journal, 73, 1896-1903.

Nørgaard L; Saudland A; Wagner J; Nielsen J P; Munck L; Engelsen S B (2000). Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. Applied Spectroscopy, 54 (3), 413-419.

Nowkandeh S M; Homaee M; Noroozi A A (2013). Mapping Soil Organic Matter Using Hyperion Images. International Journal of Agronomy and Plant Production, 4 (8), 1753-1759.

Olsen R A; Army T J; Hanway J J; Kilmer V J (1971). Fertilizer Technology and Use (2nd edition), Soil Science Society of America, Inc. Madison, Wisconsin (USA).

Pérez H; Rodríguez I; Betancourt Y; Gómez J M; Más R; Rodríguez L; Brito G; Gutiérrez J L (2001). Assessment of physical condition of CAI lands from Villa Clara province (10 briefs). National Research Institute on Sugar Cane. La Habana.

Phillips S. (2009). The role of spatial variability in nutrient management, Fall 9. International

Plant Nutrition Institute (IPNI), Georgia, USA.
Ramirez-Lopez L; Behrens T; Schmidt K; Stevens A; Demattê J A M; Scholten T (2013). The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma, 195-196, 268-279
Rowse H R; Stone D A (1980). Deep cultivation of a sandy loam, I. Effects on growth, yield and nutrient content of potatoes, broad beans, summer cabbage and red beet in 1977. Soil and Tillage Research, 1(1), 57-68.
Saeys W; Ramon H (2004). Rapid analysis of hog manure using visual and near-infrared reflectance spectroscopy, in: AgEng 2004 Conference "Engineering the future location", Leuven, Belgium, 12-16 September.
Shao Y; He Y (2011). Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. Soil Research, 49, 166-172.
Smola A J; Scholkopf B (1998). A Tutorial on Support SectorRegression; NeuroCOLT2 Technical Report Series, NC-TR-98-030; Royal Holloway College, University of London: London, U.K.
Stenberg B; Viscarra Rossel R A; Mouazen A M; Wetterlind J (2010). Visible and near infrared spectroscopy in soil science. Advances in Agronomy, 107, 163-215.
Thomasson J A; Sui R; Cox M S; Al–Rajehy A (2001). Soil reflectance sensing for determining soil properties in precision agriculture. Transactions of the ASAE, 44(6), 1445–1453
van Schöll L. and Nieuwenhuis R. (2004). Soil fertility management. 4th edn Agromisa Foundation, Wageningen, 83.
Vapnik V (1995). The Nature of Statistical Learning Theory. New York, NY: Springer-Verlag.
Vapnik V; Golowich S; Smola A (1997). Support vector method for function approximation, regression estimation, and signal processing in: Advances in Neural Information Processing Systems 9; Mozer M; Jordan M; Petsche T (Eds). MIT Press: Cambridge, MA. pp 281-287.
Volkan A; van Es H M; Akbas F; Durak A; Hively W D (2010). Visible near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. Journal of Arid Environments, 74, 229-238.
Wang S Q; Shu N; Zhang H T (2008). In-site total N content prediction of soil with Vis/NIR spectroscopy. Guang Pu Xue Yu Guang Pu Fen Xi, 28 (4), 808-812.
Wang J; Qin Q; Dong H; Chen Ch; Meng Q (2013). Study on quantitative retrieval of soil organic matter based on bare soil spectrum, in: 3rd International Conference on Multimedia Technology (ICMT 2013).

Wetterlind J; Stenberg B; Jonsson A (2008). Near infrared reflectance spectroscopy compared with soil clay and organic matter content for estimating within a field variation in N uptake in cereals. Plant and Soil, 302, 317-327.

Yang H; Kuang B; Mouazen A M (2011). Prediction of soil TN and TC at a farm-scale using VIS-NIR spectroscopy. Advanced Materials Research, 225-226 (1-2), 1258-1261.

Yu S; Yao Y; Wang D; Si H (2013). Studies on soil organic matter content mapping using EO-1 hyperion data, in: Second International Conference on Agro-Geoinformatics 2013, 439 - 443

Tables

Table 1 –Distribution of Cambisol soils and spatial variability in sugarcane yield

Municipality	CAI [a]	Area [b] (ha)	Cambisol [c] (%)	Yield Mean (t ha^{-1})	Yield Min [d] (t ha^{-1})	Yield Max [e] (t ha^{-1})	SD [f]	CV [g] (%)
Corralillo	Q. B.	17 101.80	46.12	63.10	19.30	96	19.30	32.98
Santo Domingo	G. W.	14 762.60	35.42	51.65	18.70	105.50	18.70	50.21
Santo Domingo	C. B.	6 147.10	24.12	63.56	32.10	92	32.10	31.79
Quemado de Güines	P. G. T.	15 342.50	21.40	62.03	23.20	96	23.20	35.36
Sagua La Grande	H. R.	11 214.80	10.56	71.43	24	108	24	35.82
Encrucijada	P. F.	8 176.20	5.28	64.51	22.50	96	22.50	37.74
Encrucijada	A. S.	7 945.10	25.51	76.91	39.50	97.10	39.50	20.24
Camajuaní	J. M. P.	13 782.40	44.80	70.07	22.70	102.20	22.70	28.92
Remedios	H. D.	6 424.90	26.67	50.72	24	96	24	40.98
Ranchuelo	I. A.	6 403.60	49	52.60	25.15	77.99	25.15	24.57

Number of samples (n) = 115; [a] CAI- Agroindustrial complexes; [b]Area- Total area of each CAI;
[c] Cambisol (%)- percent distribution of Cambisol soils in each area; [d] Min- minimum observed value;
[e] Max- maximum observed value; [f] SD- standard deviation in the reference data;
[g] CV- coefficient of variation of yield in one year;
Q.B.- Quintín Banderas; G.W.- George Washington; C.B- Carlos Baliño; P.G.T.- Panchito Gómez Toro;
H.R.- Héctor Rodríguez; P.F.- Perucho Figueredo; A.S.- Abel Santamaría; J.M.P.- José María Pérez;
H.D.- Heriberto Duquesne; I.A.- Ifraín Alfonso.

Table 2 – Analytical methods for determination of soil fertility parameters.

Soil fertility parameter	Conventional method	Extraction	Determination
Organic matter	Walkley-Black	Digestion $K_2Cr_2O_7 + H_2SO_4$	Colorimetry
Available potassium	Oniani	0.1 *M* H_2SO_4	Flame photometry
Available phosphorus	Olsen	0.5 *M* $NaHCO_3$	Colorimetry
Available phosphorus	Oniani	0.1 *M* H_2SO_4	Colorimetry

Table 3 – Guidelines for assessing the calibration accuracy.

Calibration accuracy	R^2 [a]	RPD [b]	RER [c]	Source
Excellent	$R^2 > 0.95$	RPD >4	RER > 20	Malley et al. (2004)
Successful	$0.90 \leq R^2 \leq 0.95$	$3 \leq$ RPD ≤ 4	$15 \leq$ RER ≤ 20	Malley et al. (2004)
Moderately successful	$0.80 \leq R^2 < 0.90$	$2.25 \leq$ RPD < 3	$10 \leq$ RER < 15	Malley et al. (2004)
Moderately useful	$0.70 \leq R^2 < 0.80$	$1.75 \leq$ RPD < 2.25	$8 \leq$ RER < 10	Malley et al. (2004)
Less reliable	$R^2 < 0.70$	RPD < 1.75	RER < 8	Nduwamungu et al. (2009)

[a] R^2- coefficient of determination; [b] RPD- ratio of performance deviation; [c] RER- ratio of error range

Table 4 – Basic statistics for concentrations of soil fertility parameters at landscape level.

Soil fertility parameter	Mean	Min [a]	Max [b]	SD [c] (±)	SEL [d]	CV [e] (%)
OM	2.95	2.23	3.98	0.42	0.000	14.08
K_2O	19.48	9.06	35.36	7.11	0.569	36.49
Olsen P	2.55	1.68	4.10	0.58	0.001	22.87
Oniani P	12.00	5.10	25.02	5.36	0.194	44.62

OM in %; K_2O in mg K_2O 100 g^{-1} d.s, Olsen P and Oniani P in mg P 100 g^{-1} d.s
[a] Min- minimum observed value; [b] Max- maximum observed value; [c] SD- standard deviation in the reference data;
[d] SD- Standard error of laboratory; [e] CV- coefficient of variation

Table 5 – Basic statistics for concentrations of soil fertility parameters within a field.

Soil fertility parameter	Mean	Min [a]	Max [b]	SD [c] (±)	SEL [d]	CV [e] (%)
OM	3.04	2.54	3.56	0.37	0.001	12.16
K_2O	16.55	7.89	30.58	6.84	0.995	37.68
Olsen P	2.39	1.75	3.15	0.36	0.005	15.27
Oniani P	13.90	6.55	22.84	4.39	0.240	31.70

OM in %; K_2O in mg K_2O 100 g^{-1} d.s (dry soil), Olsen P and Oniani P in mg P 100 g^{-1} d.s.
[a] Min- minimum observed value; [b] Max- maximum observed value; [c] SD- standard deviation in the reference data;
[d] SD- Standard error of laboratory; [e] CV- coefficient of variation

Table 6 – Pearson correlation coefficients among soil fertility parameters at landscape level.

	OM	K_2O	Olsen P	Oniani P
OM	1.00			
K_2O	0.78*	1.00		
Olsen P	0.89*	0.68*	1.00	
Oniani P	0.88*	0.76*	0.83*	1.00

*Significant at the level of <0.01

Table 7 – Pearson correlation coefficients among soil fertility parameters within a field.

	OM	K_2O	Olsen P	Oniani P
OM	1.00			
K_2O	0.77*	1.00		
Olsen P	0.85*	0.65*	1.00	
Oniani P	0.83*	0.72*	0.78*	1.00

*Significant at the level of <0.01

Table 8 – Comparison between the best predictive models with others calibrated at landscape level.

Soil fertility parameter	Regression model	R^2 C [a]	R^2 CV [b]	R^2 P [c]	RMSEC [d]	RMSECV [e]	RMSEP [f]	RPD [g]	RER [h]	C Bias [i]	CV Bias [j]	P Bias [k]
OM	PLS	0.93	0.93	0.90	0.10	0.10	0.14	3.29	12.32	-0.00	-0.00	0.01
	LWR	**0.96**	**0.95**	**0.93**	**0.07**	**0.08**	**0.12**	**3.96**	**14.80**	**0.02**	**0.02**	**0.03**
	SVM	0.96	0.95	0.91	0.08	0.08	0.14	3.32	12.41	-0.00	-0.00	0.02
K_2O	PLS	0.77	0.75	0.77	3.00	3.12	3.62	2.08	7.26	-1.06	-0.01	0.03
	LWR	**0.83**	**0.79**	**0.79**	**2.72**	**3.07**	**3.47**	**2.79**	**9.74**	**0.48**	**0.51**	**2.18**
	SVM	0.83	0.78	0.78	2.66	3.02	3.49	2.17	7.57	-0.15	0.04	0.33
Olsen P	PLS	0.71	0.69	0.69	0.29	0.31	0.35	1.82	6.21	-0.00	-0.00	-0.05
	LWR	0.83	0.78	0.81	0.23	0.26	0.27	2.35	8.01	0.03	0.04	-0.03
	SVM	**0.83**	**0.81**	**0.81**	**0.22**	**0.24**	**0.27**	**2.37**	**8.10**	**0.00**	**-0.00**	**-0.05**
Oniani P	PLS	0.75	0.73	0.64	2.57	2.65	3.54	1.60	5.56	-1.06	-0.01	0.39
	LWR	0.80	0.74	0.65	2.33	2.62	3.61	1.61	5.61	0.37	0.40	0.93
	SVM	**0.79**	**0.73**	**0.65**	**2.37**	**2.65**	**3.31**	**1.70**	**5.91**	**-0.21**	**-0.12**	**0.03**

[a] R^2 C- coefficient of determination of calibration; [b] R^2 CV- coefficient of determination of cross validation; [c] R^2 P- coefficient of determination of prediction
[d] RMSEC- root mean square error of calibration; [e] RMSECV- root mean square error of cross validation; [f] RMSEP- root mean square error of prediction;
[g] RPD- ratio of performance deviation; [h] RER- ratio of error range;
[i] C Bias- systematic deviation of calibration; [j] CV Bias- systematic deviation of cross validation; [k] P Bias- systematic deviation of prediction
Best regression model based on the R^2, RMSECV, RMSEP, RPD, RER

Table 9 – Calibration and prediction statistics of the calibration models within a field.

Soil fertility parameter	Regression model	R^2 C	R^2 CV	R^2 P	RMSEC [d]	RMSECV [e]	RMSEP [f]	RPD [g]	RER [h]	C Bias [i]	CV Bias [j]	P Bias [k]
OM	PLS	0.96	0.95	0.92	0.08	0.09	0.14	3.22	8.88	-0.00	-0.00	-0.08
	LWR	0.96	0.95	0.92	0.07	0.09	0.14	3.45	9.51	0.01	0.02	-0.09
	SVM	**0.97**	**0.97**	**0.92**	**0.07**	**0.07**	**0.14**	**3.45**	**9.51**	**-0.00**	**-0.00**	**-0.09**
K_2O	PLS	0.74	0.73	0.61	3.53	3.63	5.88	1.56	5.66	-1.07	0.02	4.30
	LWR	**0.83**	**0.76**	**0.63**	**2.90**	**3.43**	**5.13**	**1.61**	**5.87**	**0.33**	**0.43**	**3.37**
	SVM	0.80	0.75	0.63	3.08	3.53	5.43	1.52	5.53	-0.18	-0.07	3.56
Olsen P	PLS	0.72	0.70	0.68	0.30	0.31	0.34	1.29	4.97	0	-0.00	0.19
	LWR	0.89	0.82	0.77	0.19	0.25	0.34	1.46	5.59	0.02	0.02	-0.23
	SVM	**0.87**	**0.83**	**0.83**	**0.20**	**0.23**	**0.27**	**2.95**	**11.32**	**-0.00**	**-0.00**	**0.24**
Oniani P	PLS	0.74	0.73	0.70	2.71	2.77	2.52	1.76	6.50	-1.07	0.01	0.26
	LWR	**0.82**	**0.77**	**0.72**	**2.26**	**2.58**	**2.32**	**1.90**	**7.02**	**0.33**	**0.42**	**-0.06**
	SVM	0.82	0.79	0.71	2.44	2.36	2.48	1.78	6.58	-0.06	-0.01	0.16

[a] R^2 C- coefficient of determination of calibration; [b] R^2 CV- coefficient of determination of cross validation; [c] R^2 P- coefficient of determination of prediction

[d] RMSEC- root mean square error of calibration; [e] RMSECV- root mean square error of cross validation; [f] RMSEP- root mean square error of prediction;

[g] RPD- ratio of performance deviation; [h] RER- ratio of error range;

[i] C Bias- systematic deviation of calibration; [j] CV Bias- systematic deviation of cross validation; [k] P Bias- systematic deviation of prediction

Best regression model based on the R^2, RMSECV, RMSEP, RPD, RER

Table 10 – Comparison between the predictive regression models at landscape level based on the selected wavelength intervals.

Soil fertility parameter	Regression model	R^2C [a]	R^2CV [b]	R^2P [c]	RMSEC [d]	RMSECV [e]	RMSEP [f]	RPD [g]	RER [h]	C Bias [i]	CV Bias [j]	P Bias [k]
OM	PLS	0.95	0.95	0.96	0.08	0.09	0.09	5.11	19.11	-0.00	-0.00	0.00
	LWR	**0.98**	**0.98**	**0.97**	**0.06**	**0.07**	**0.08**	**5.80**	**21.67**	**0.01**	**0.01**	**0.01**
	SVM	0.99	0.97	0.96	0.05	0.07	0.11	4.18	15.64	-0.00	-0.00	0.00
K_2O	PLS	0.79	0.77	0.81	2.98	3.07	3.30	2.29	7.99	-0.00	0.00	0.26
	LWR	**0.84**	**0.80**	**0.82**	**2.56**	**2.92**	**3.25**	**2.32**	**8.09**	**0.16**	**0.13**	**0.04**
	SVM	0.84	0.80	0.82	2.60	2.92	3.30	2.28	7.97	-0.01	-0.02	0.01
Olsen P	PLS	0.69	0.68	0.70	0.29	0.31	0.34	1.87	6.39	-0.00	-0.00	-0.05
	LWR	0.86	0.76	0.87	0.21	0.26	0.23	2.78	9.49	0.01	0.02	-0.04
	SVM	**0.83**	**0.80**	**0.89**	**0.23**	**0.24**	**0.22**	**3.14**	**10.71**	**-0.02**	**-0.02**	**-0.09**
Oniani P	PLS	0.77	0.76	0.69	2.47	2.54	3.29	1.73	6.00	-0.00	0.02	0.43
	LWR	0.81	0.77	0.68	2.28	2.53	3.38	1.70	5.91	0.28	0.33	0.68
	SVM	**0.79**	**0.76**	**0.71**	**2.36**	**2.53**	**3.04**	**1.86**	**6.45**	**0.01**	**-0.00**	**0.18**

[a] R^2C- coefficient of determination of calibration; [b] R^2CV- coefficient of determination of cross validation; [c] R^2P- coefficient of determination of prediction

[d] RMSEC- root mean square error of calibration; [e] RMSECV- root mean square error of cross validation; [f] RMSEP- root mean square error of prediction;

[g] RPD- ratio of performance deviation; [h] RER- ratio of error range;

[i] C Bias- systematic deviation of calibration; [j] CV Bias- systematic deviation of cross validation; [k] P Bias- systematic deviation of prediction

Best regression model based on the R^2, RMSECV, RMSEP, RPD, RER

Table 11 – Comparison between the predictive regression models within a field based on the selected wavelength intervals.

Soil fertility parameter	Regression model	R^2 C	R^2 CV	R^2 P	RMSEC[d]	RMSECV[e]	RMSEP[f]	RPD[g]	RER[h]	C Bias[i]	CV Bias[j]	P Bias[k]
OM	PLS	0.95	0.95	0.92	0.08	0.08	0.15	3.31	9.12	-0.00	-0.00	-0.10
	LWR	0.98	0.97	0.92	0.05	0.07	0.18	3.72	10.25	0.01	0.01	-0.15
	SVM	**0.98**	**0.96**	**0.92**	**0.05**	**0.06**	**0.18**	**3.72**	**10.25**	**-0.00**	**0.07**	**-0.14**
K_2O	PLS	0.76	0.73	0.63	3.36	3.53	5.50	1.56	5.69	-0.00	-0.00	3.79
	LWR	**0.83**	**0.75**	**0.65**	**2.79**	**3.47**	**4.47**	**2.88**	**10.47**	**0.17**	**0.30**	**3.91**
	SVM	0.80	0.76	0.64	3.04	3.33	4.53	1.65	5.99	-0.35	-0.19	2.48
Olsen P	PLS	0.74	0.73	0.69	0.29	0.29	0.30	1.29	4.95	-0.00	0.00	0.10
	LWR	0.87	0.83	0.82	0.21	0.24	0.25	1.63	6.24	0.02	0.03	-0.11
	SVM	**0.86**	**0.82**	**0.84**	**0.21**	**0.24**	**0.20**	**2.40**	**9.21**	**0.01**	**-0.00**	**0.13**
Oniani P	PLS	0.79	0.78	0.73	2.42	2.50	2.71	1.71	6.34	-0.00	-0.01	-0.86
	LWR	0.84	0.79	0.73	2.12	2.44	2.69	2.08	7.70	0.19	0.19	-1.66
	SVM	**0.83**	**0.81**	**0.72**	**2.20**	**2.34**	**2.54**	**1.90**	**7.02**	**-0.08**	**-0.06**	**-1.03**

[a] R^2 C- coefficient of determination of calibration; [b] R^2 CV- coefficient of determination of cross validation; [c] R^2 P- coefficient of determination of prediction
[d] RMSEC- root mean square error of calibration; [e] RMSECV- root mean square error of cross validation; [f] RMSEP- root mean square error of prediction;
[g] RPD- ratio of performance deviation; [h] RER- ratio of error range;
[i] C Bias- systematic deviation of calibration; [j] CV Bias- systematic deviation of cross validation; [k] P Bias- systematic deviation of prediction
Best regression model based on the R^2, RMSECV, RMSEP, RPD, RER

Figures

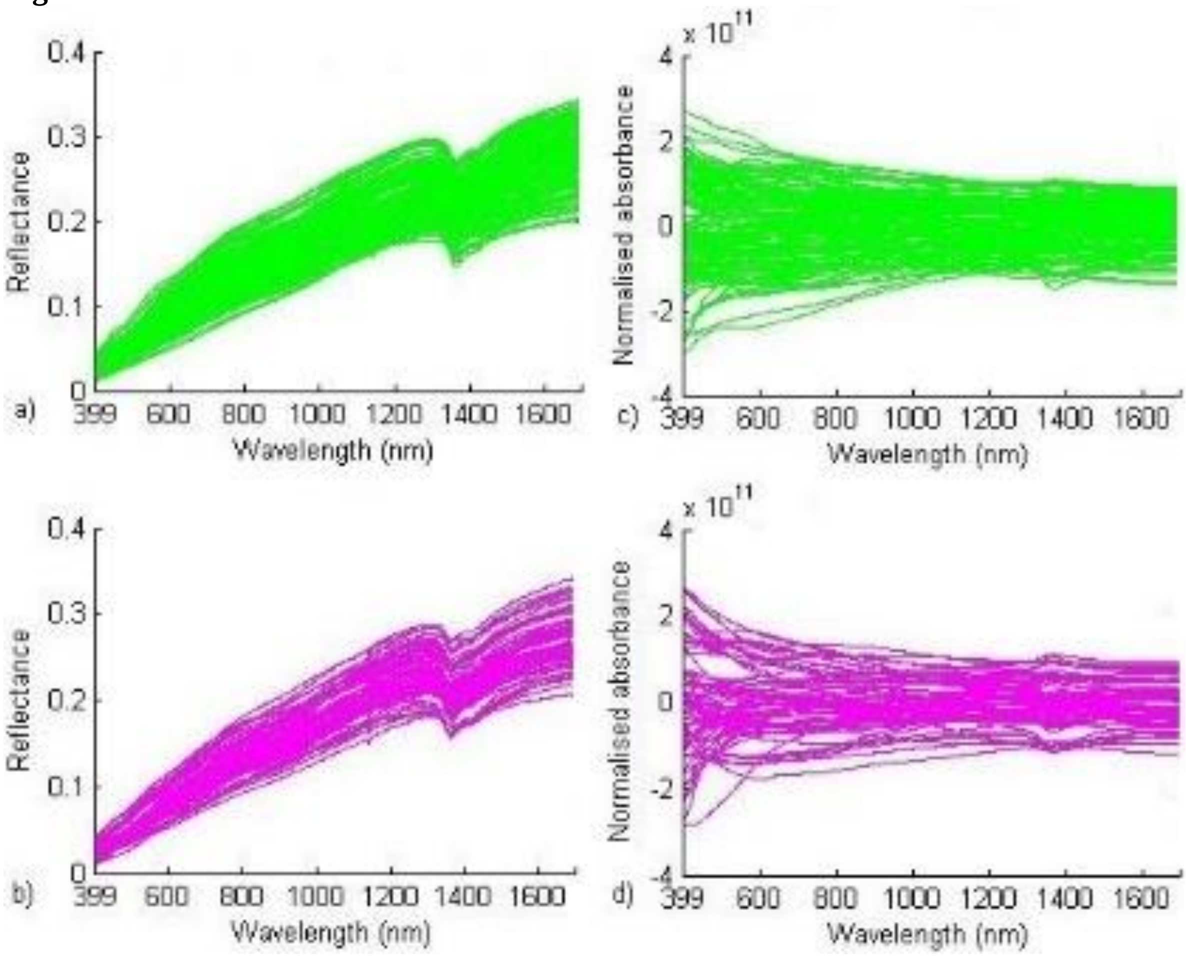


Fig. 1 – Soil spectral reflectance before and after preprocessing for measuring the average fertility at landscape level
(a) Raw data of calibration, (b) Preprocessed data of calibration, (c) Raw data of validation, (d) Preprocessed data of validation.

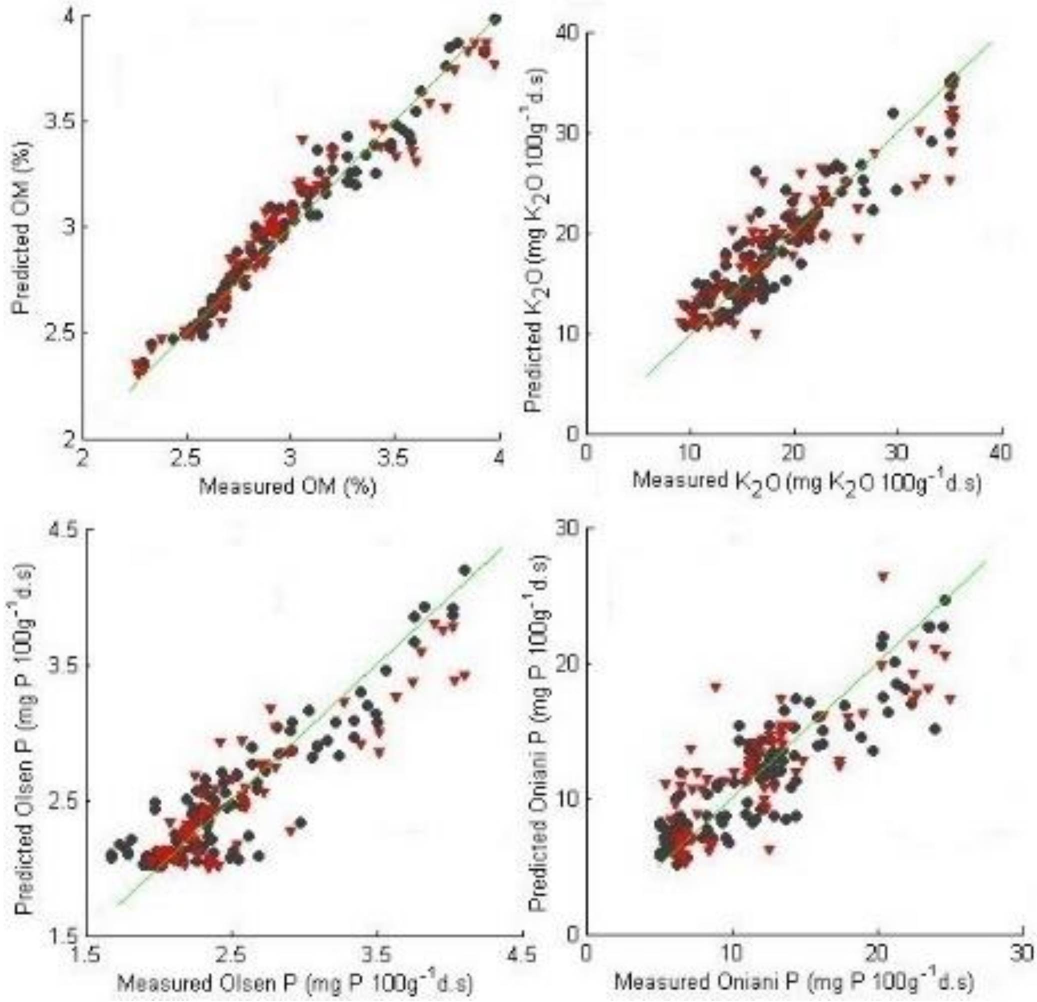


Fig. 2 – Vis/NIR predicted versus chemical conventional analyses of soil fertility parameters at landscape level; OM (LWR), K_2O (LWR), Olsen P (SVM), Oniani P (SVM).

R^2- coefficient of determination, RMSEP- root mean square error of prediction,

•- cross-validation, ▼- test set prediction

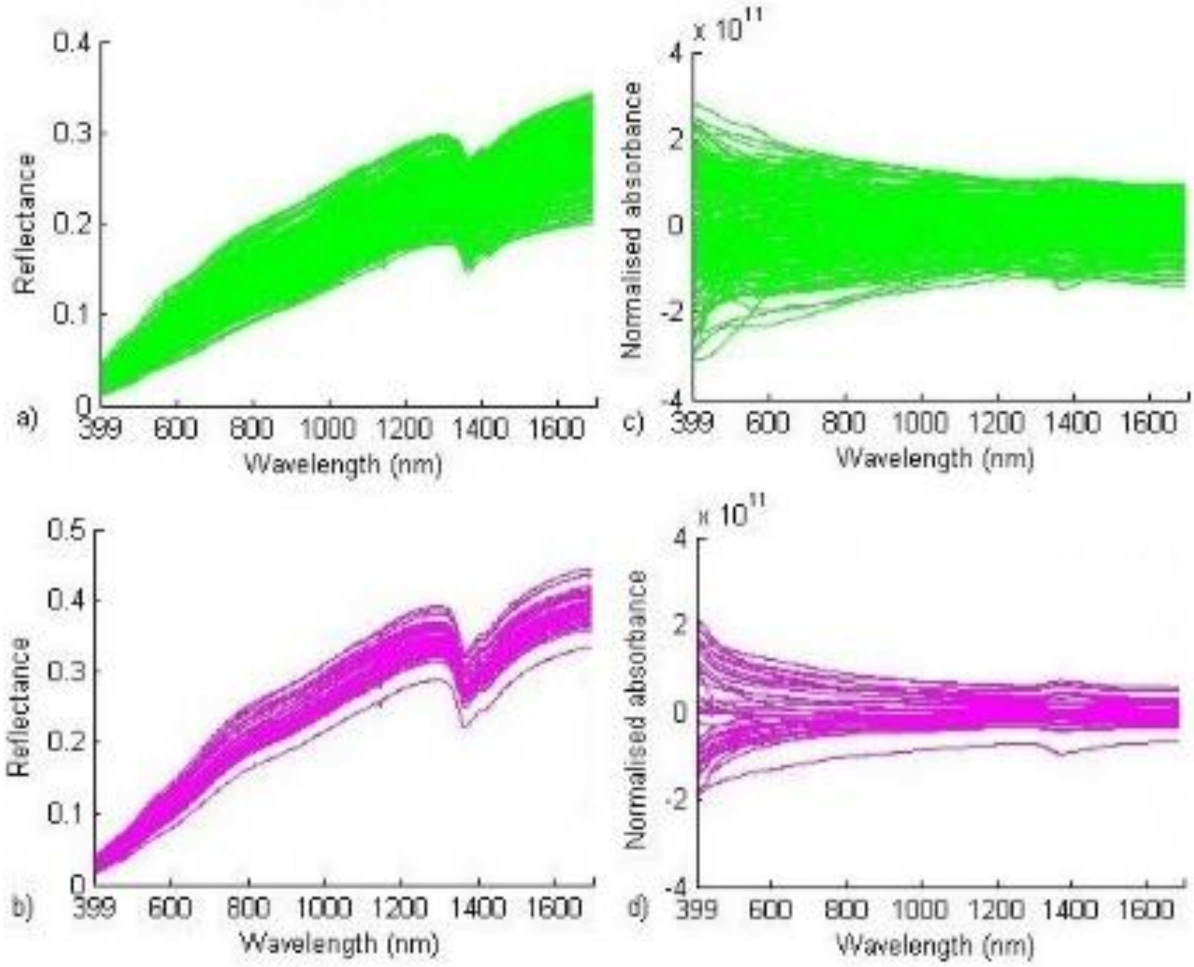


Fig. 3 – Soil spectral reflectance before and after preprocessing for measuring the average fertility within a field (a) Raw data of calibration, (b) Preprocessed data of calibration, (c) Raw data of validation (d) Preprocessed data of validation.

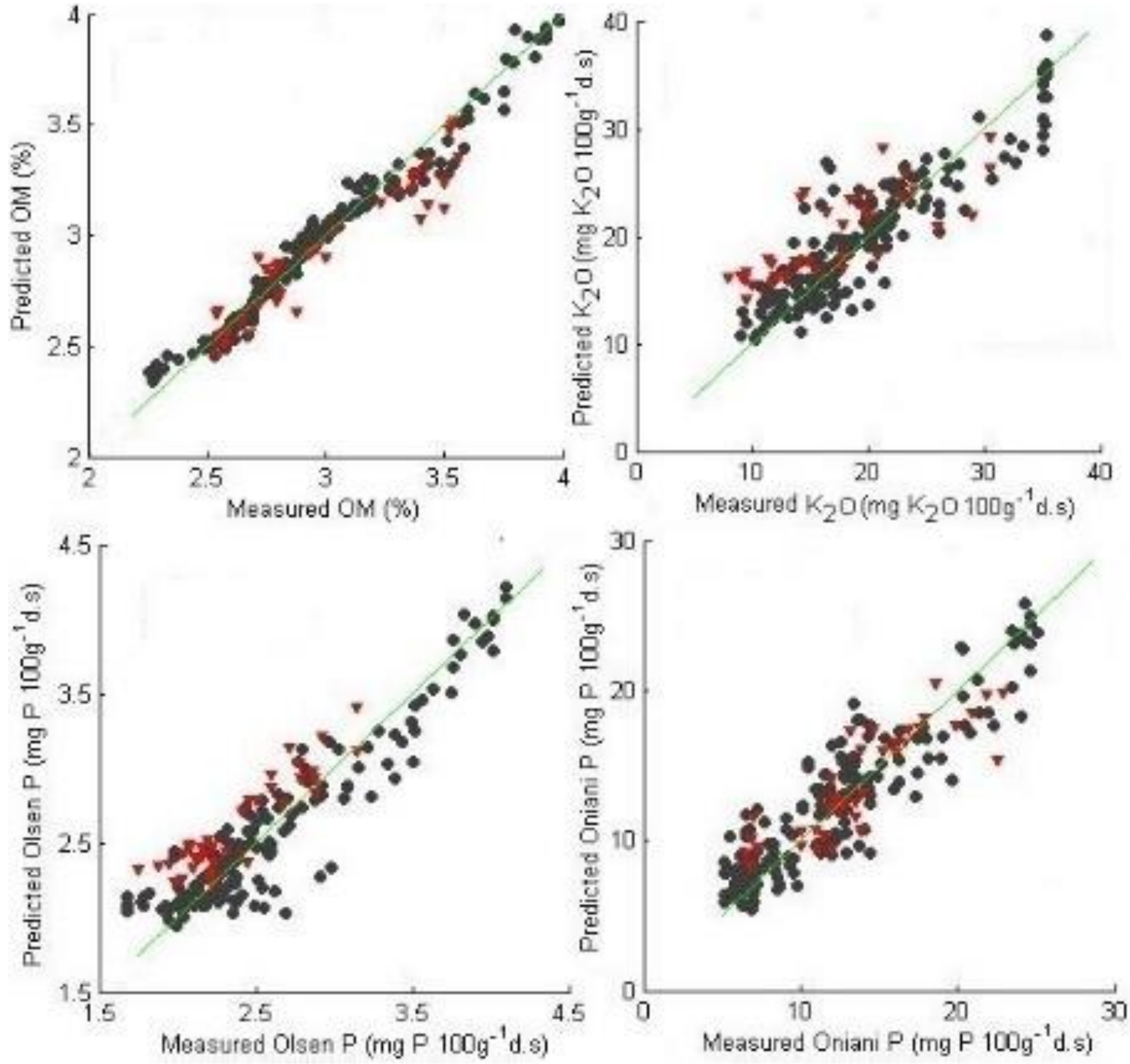


Fig. 4 – Vis/NIR predicted versus chemical conventional analyses of soil fertility parameters within a field; OM (SVM), K_2O (LWR), Olsen P (SVM), Oniani P (LWR).
R^2- coefficient of determination, RMSEP- root mean square error of prediction,
•- cross-validation, ▼- test set prediction

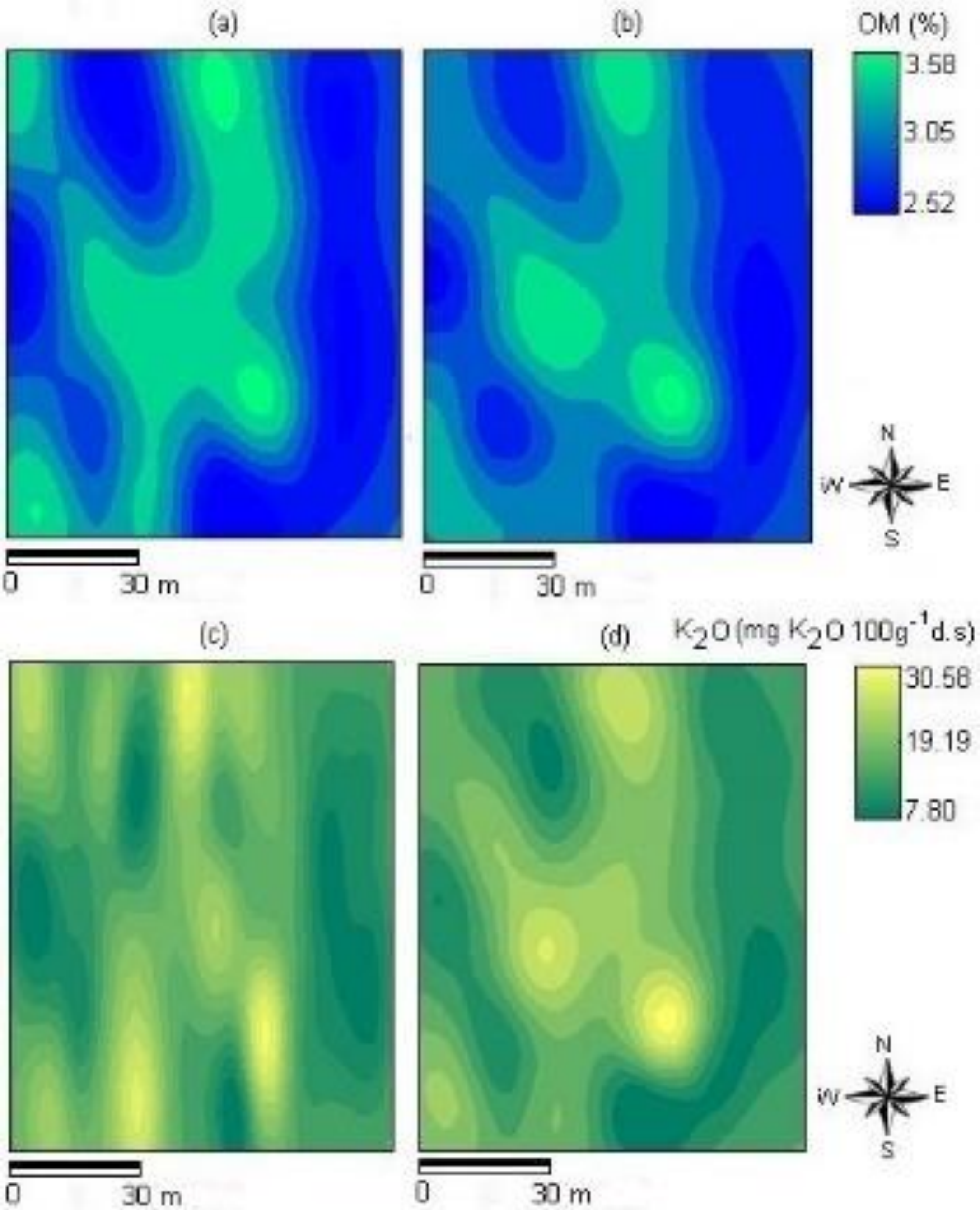


Fig. 5 – Comparison of wet chemically measured (left) and Vis/NIR predicted (right) soil fertility parameters within a field.

(a) measured OM, (b) predicted OM, (c) measured K_2O, (d) predicted K_2O).

Coefficients of variation (CV) for wet chemically measured values (OM- 12.16%, K_2O- 37.68%)

Coefficients of variation (CV) for Vis/NIR predicted values (OM- 12.34%, K_2O- 19.88%)

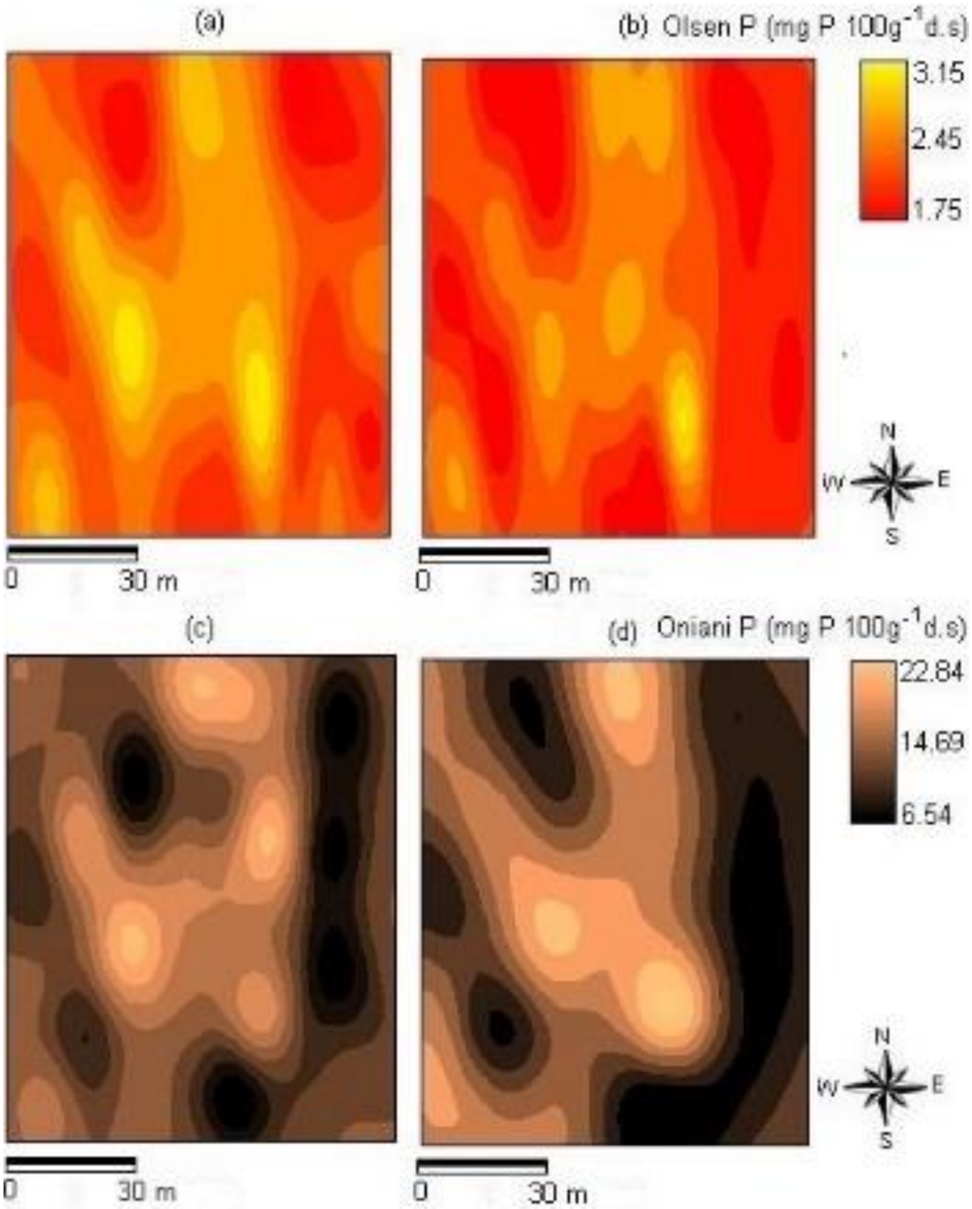
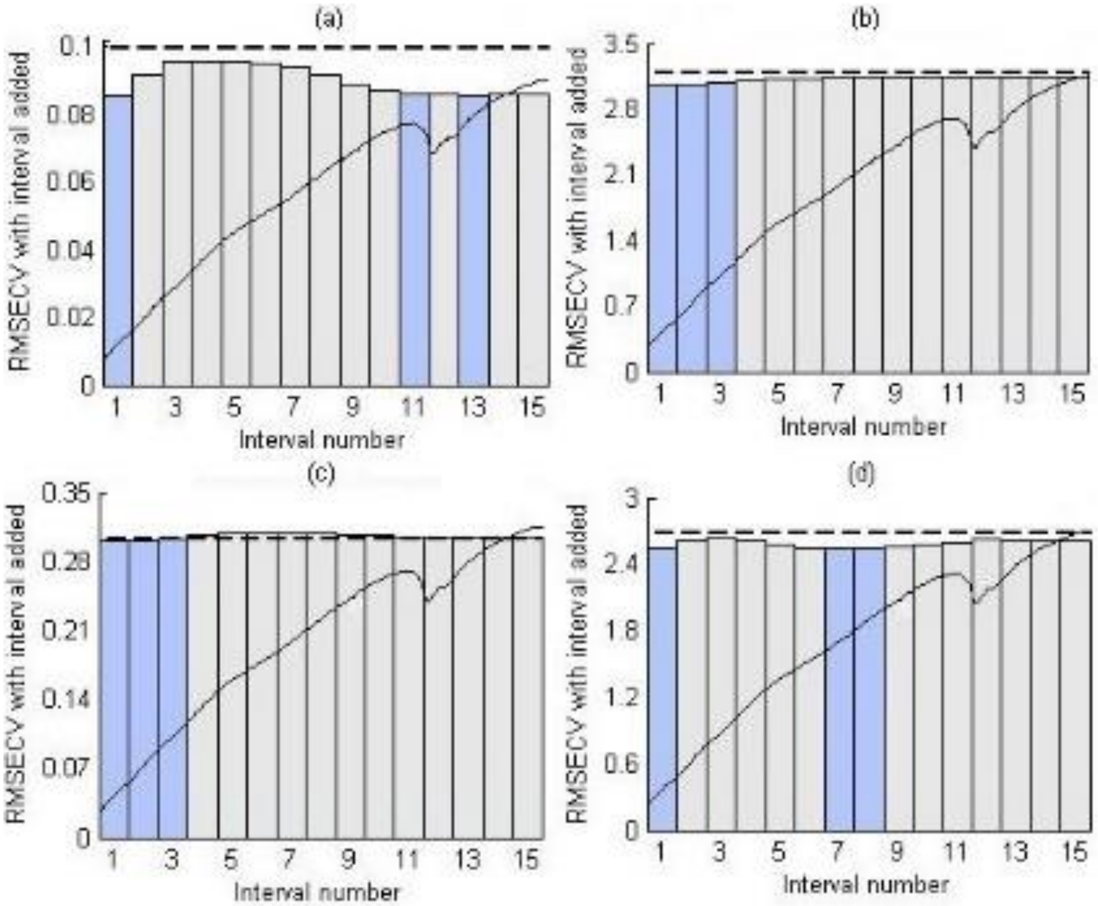


Fig. 6 – Comparison of wet chemically measured (left) and Vis/NIR predicted (right) soil fertility parameters within a field.

(a) measured Olsen P, (b) predicted Olsen P, (c) measured Oniani P, (d) predicted Oniani P.

Coefficients of variation (CV) for wet chemically measured values (Olsen P- 15.27%, Oniani P- 31.70%)

Coefficients of variation (CV) for Vis/NIR predicted values (Olsen P- 12.34%, Oniani P- .70%)



80
81 Fig. 7 – Spectral region selected (dark color) by Forward Interval PLS model (iPLS) at landscape level.

82 Intervals selected:

83 (a) OM- [1 (399 – 489 nm), 11 (1320 – 1402 nm), 13 (1480 – 1555 nm)]
84 (b) K_2O- [1 (399 – 489 nm), 2 (489 – 580 nm), 3 (580 – 674 nm)]
85 (c) Olsen P- [1 (399 – 489 nm), 2 (489 – 580 nm), 3 (580 – 674 nm)]
86 (d) Oniani P- [1 (399 – 489 nm), 7 (952 – 1050 nm), 8 (1050 – 1145 nm)]
87
88

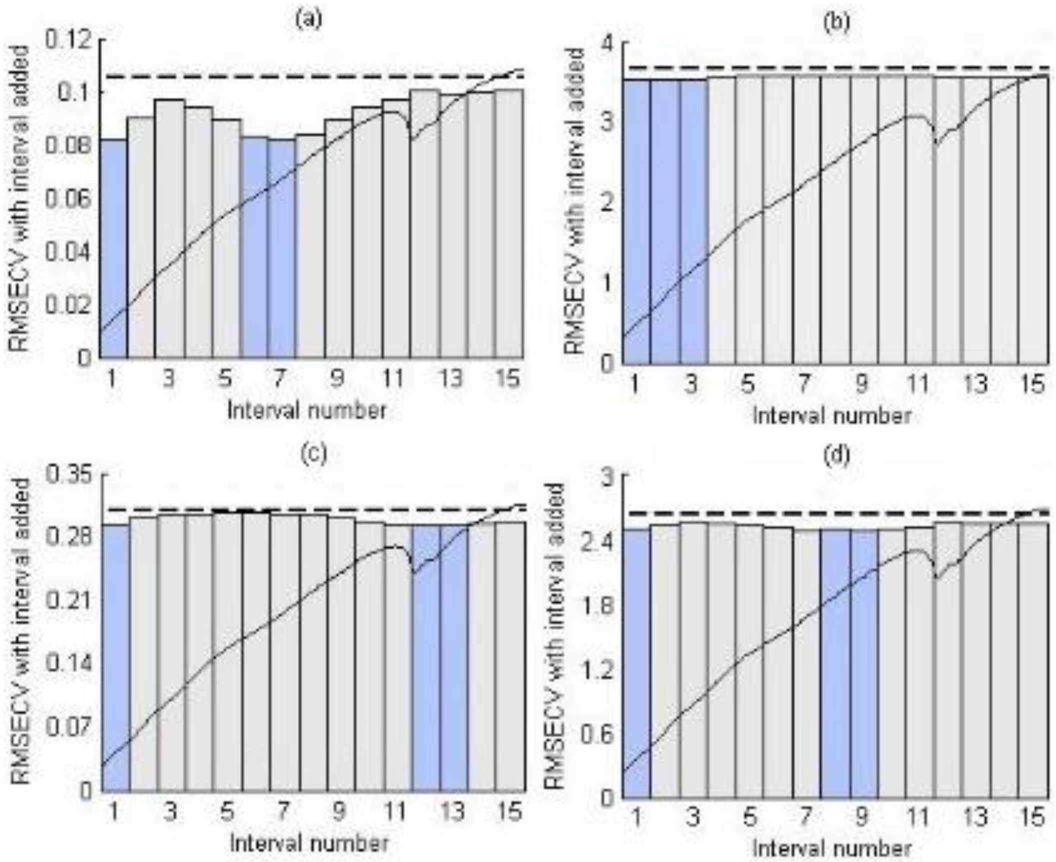


Fig. 8 – Spectral region selected (dark color) by Forward Interval PLS model (iPLS) within a field.

Intervals selected:

(a) OM- [1 (399 – 489 nm), 6 (860 – 952 nm), 7 (952 – 1050 nm)]
(b) K_2O- [1 (399 – 489 nm), 2 (489 – 580 nm), 3 (580 – 674 nm)]
(c) Olsen P- [1 (399 – 489 nm), 12 (1402 – 1480 nm), 13 (1480 – 1555 nm)]
(d) Oniani P- [1 (399 – 489 nm), 8 (1050 – 1145 nm), 9 (1145 – 1235 nm)]