

Capability of Visible-Near Infrared Spectroscopy in Estimating Soils Carbon, Potassium and Phosphorus

Bakamba dite Djeneba Sacko¹, Souleymane Sanogo^{1*}, Hamidou Konare², Abdramane Ba¹, Tiéman Diakite²

¹Laboratoire d'Optique de Spectroscopie et des Sciences Atmosphériques (LOSSA), FST, Université des Sciences Techniques et des Technologies de Bamako (USTTB), Bamako, Mali

²Laboratoire Sol-Eau-Plantes (LSEP) de l'Institut d'Economie Rural (IER) de Sotuba, Bamako, Mali

Email: *soul@lossa-mali-edu.org

How to cite this paper: dite Djeneba Sacko, B., Sanogo, S., Konare, H., Ba, A. and Diakite, T. (2018) Capability of Visible-Near Infrared Spectroscopy in Estimating Soils Carbon, Potassium and Phosphorus. *Optics and Photonics Journal*, 8, 123-134.

<https://doi.org/10.4236/opj.2018.85012>

Received: February 14, 2018

Accepted: May 28, 2018

Published: May 31, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The spectroscopy technique has many advantages over conventional analytical methods since it is fast and easy to implement and with no use of chemical extractants. The objective of this study is to quantify soil total Carbon (C), available Phosphorus (P) and exchangeable potassium (K) using VIS-NIR reflectance spectroscopy. A total of 877 soils samples were collected in various agricultural fields in Mali. Multivariate analysis was applied to the recorded soils spectra to estimate the soil chemical properties. Results reveal the over performance of the Principal Component Regression (PCR) compared to the Partial Least Square Regression (PLSR). For coefficient of determination (R²), PLSR accounts for 0.29, 0.42 and 0.57; while the PCR gave 0.17, 0.34 and 0.50, respectively for C, P and K. Nevertheless, this study demonstrates the potential of the VIS-NIR reflectance spectroscopy in analyzing the soils chemical properties.

Keywords

VIS-NIR Spectroscopy, Soils Chemical Properties, Principal Component Regression (PCR), Partial Least Square Regression (PLSR)

1. Introduction

For the large majority of countries, agriculture is an area of particular interest, in social, economic and environmental points of view [1]. It plays a key role in Malian economy and the country food security. Quantification of soils nutrient levels available to plants contributes to the success of productive cropping systems

and healthy environment [2]. In Mali, chemical methods are widely used to make measure soil physical and chemical properties. However, these methods require the use of chemical extractants which are expensive and harmful to human and his environment. In this context, the search for an alternative solution to replace or reduce conventional methods for soil property analyses becomes necessary.

Advantages of spectroscopy methods are justified for several reasons. For example, sample preparation involves only drying and grinding, thus the sample chemical properties are not affected by the analysis. Also, the measurement is fast and several soil properties can be estimated from a single scan. Moreover, the technique can be performed in the laboratory or directly *in situ*.

First publications on the potential of VIS-NIR spectroscopy for soil analysis appeared in the early 1990 [3] [4]. Since this period, many works have been carried out on the use of this technique, especially during the last decades [5]. Particular emphasis has been placed on the classical soil properties such as soil organic matter (SOM), clay content, mineralogy, chemical nutrients, structure and microbial activity [5]. According to [6], calibrations for total organic carbon as well as total soil nitrogen are the most likely to be successful. Previous findings indicated a coefficient of determination ranging from 0.23 to 0.92 for P and 0.11 to 0.55 for exchangeable K [7] [8] [9]. However, predictions for P and K by spectroscopy remained unacceptable [10] [11].

Regarding the development of spectroscopic methods, few studies have been done in some African countries like Mali. Other studies [12] used soil samples from across the Lake Victoria basin of Kenya to investigate the potential of near infrared diffuse reflectance spectroscopy to estimate some properties. However, they suggested further calibrations using more diverse soil types and testing alternative infrared diffuse reflectance based methods.

The interest of the VIS-PIR spectroscopy method is justified for several reasons. For instance, sample preparation involves only drying and grinding, the sample is not affected by the analysis in any way, also no chemicals (environmental hazard) are required. In addition, the measurement is fast because it takes only a few seconds and several soil properties can be estimated simultaneously from a single scan. Moreover, the technique can be done in the laboratory or directly *in situ*.

A network of infrared spectroscopy laboratories is supported by the World Agroforestry Center (ICRAF) in national institutions in Africa, currently in Cote d'Ivoire, Kenya, Malawi, Mali, Mozambique, Nigeria, and Tanzania. Despite all this effort the spectroscopy method is poorly used for soil analysis in different research institutions. The present study contributes to document and to demonstrate the potential of the diffuse reflectance spectroscopy to estimate some soil properties in Mali.

The general objective of this study is to evaluate the performance of VIS-NIR spectroscopy in comparing the estimates of two regression models, namely Prin-

Principal Component Regression (PCR) and Partial Least Squared Regression (PLSR) for the determination of total carbon (C), available phosphorus (P) and exchangeable potassium (K).

2. Material and Methods

2.1. Sample Collection

Soil sampling and preparation were carried out by the soil laboratory “Laboratoire Sol-Eau-Plante” in Mali. Areas covered by the study are the administrative regions of Koulikoro and Sikasso where 755 and 122 samples, were collected respectively. The soil samples were collected from 0 - 10 cm depth. **Figure 1** shows the geographical characteristics of each sampling site.

2.2. Measurements of Soil Reference Data

Soil reference data measurements were carried out in the soil laboratory “Laboratoire Sol-Eau-Plante” (LSEP) using standard laboratory methods. Soil samples were priority air dried, crushed and sieved to 2 mm. The total carbon was measured by an automatic titrator using the modified Anne method which is the oxidation of soil carbon by potassium dichromate. The available phosphorus of the soil was extracted with a combined solution of 0.1 M HCl and 0.03 M NH₄F and the measurements were made using an ultraviolet (UV) spectrophotometer or a colorimeter. For the exchangeable K, the soil was leached with a 1 M ammonium acetate solution at pH 7. It was determined directly in the ammonium acetate percolate using a flame photometer.

2.3. Sample Selection Method

The 877 soils samples were partitioned into two sub-samples constituting sets of calibration and validation. The selection was made in the way that all sampling sites are represented in both calibration and validation sets. Approximately 2/3 of samples from each sampling site are selected to form the calibration sample set (587 samples) and the remaining 1/3 is used to form the validation sample set (290 samples).

2.4. Spectral Measurements

Spectral measurements and their processing were carried out at the Laboratory of Optics, Spectroscopy and Atmospheric Sciences (LOSSA). These measurements consist of recording the soils reflectance over the wavelength range of 342 - 1060 nm. Soils samples were priority air dried and crushed to pass a 2-mm sieve. A Miniature Fiber Optic Spectrometer working on UV-VIS-NIR spectral range (BLUE-Wave Miniature Fiber Optic Spectrometers for UV-VIS-NIR & OEM, StellarNet Inc.) was used to perform the spectral measurements. The spectrometer is connected to a PC on which the *Spectra Wiz* software is installed for controlling the data acquisition. The samples were scanned using a halogen tungsten

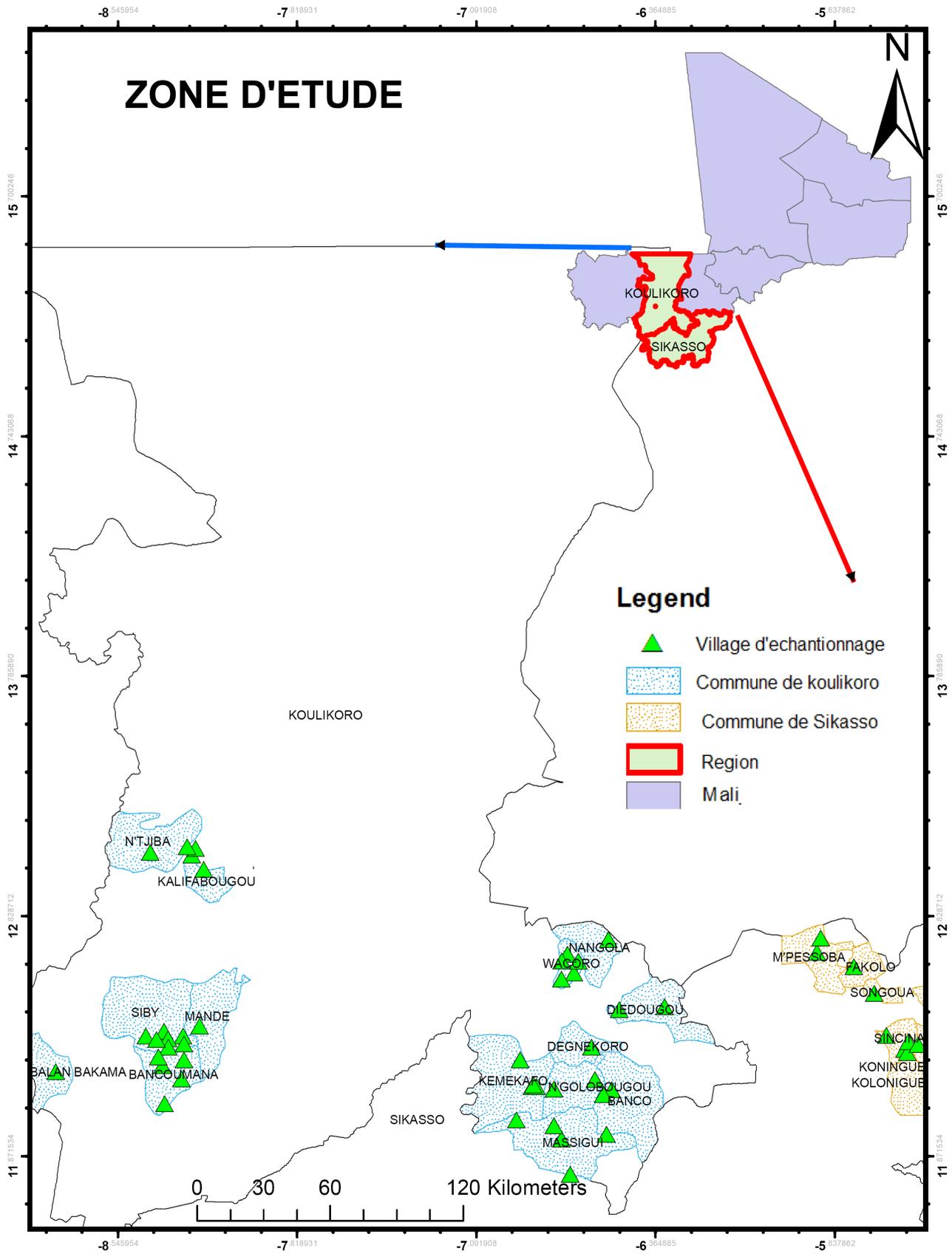


Figure 1. Maps representing the distribution of soil sampling sites in Mali.

SL1 lamp source manufactured by StellarNet Inc. This light source has a wide spectral Range for 350 - 2500 nm-effective for color, reflectance, transmittance, and absorbance measurements. The Y-shaped optical fiber was used to transport light from the source to the sample and from sample to the spectrometer (**Figure 2(a)**) with the same incident angle of illumination.

2.5. Pre-Treatment of Spectra

Before being used, the raw spectral data are undergoing various pretreatments. The most common strategy for pre-processing spectra is to submit the raw data to one or more mathematical transformations intended to make them suitable for modeling.

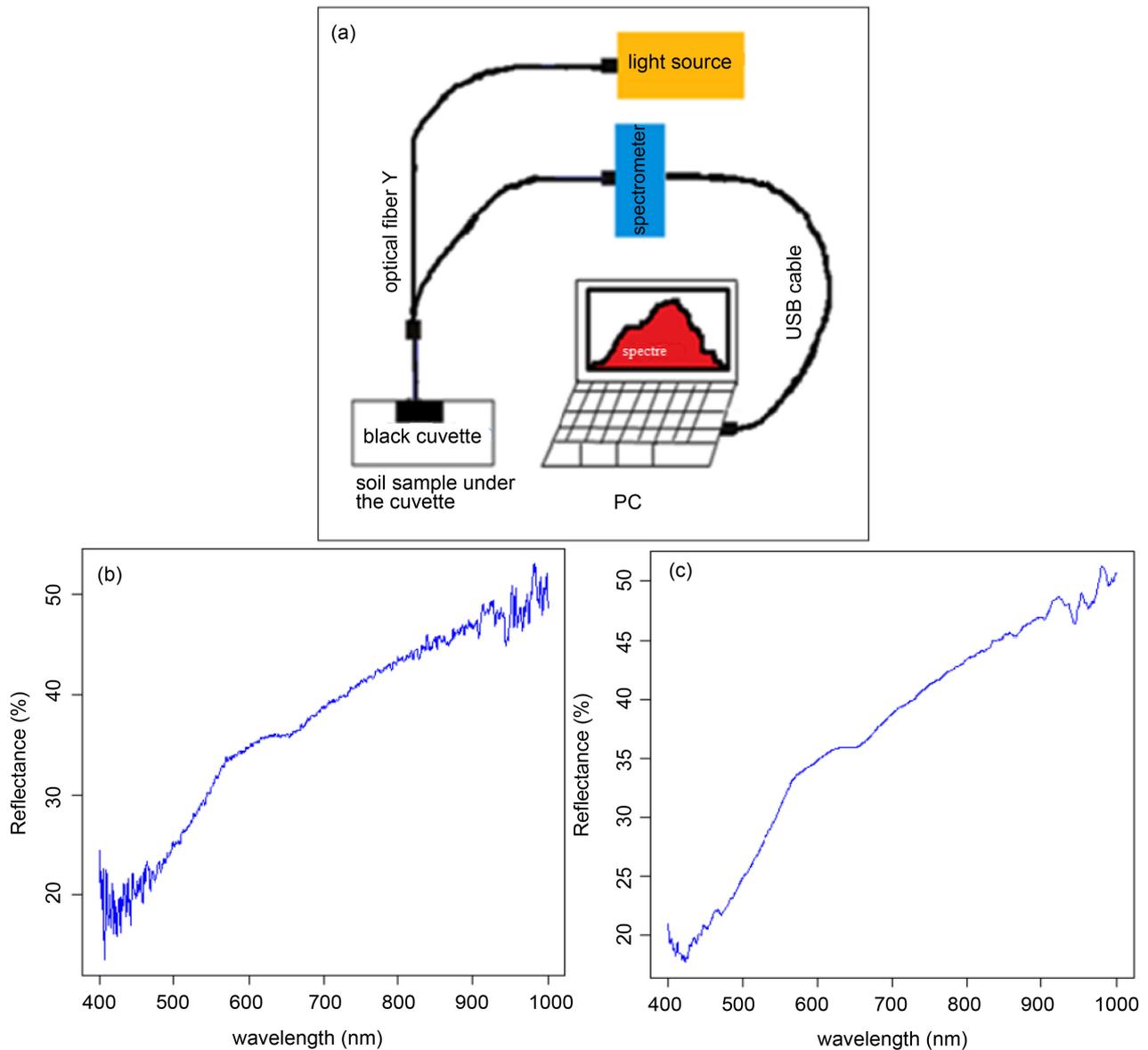


Figure 2. (a) Scheme describing the spectroscopic equipment used for scanning soil samples and spectrum acquisition; (b) Raw spectrum; (c) filtered spectrum.

Sample spectra were filtered using the *RunMean* function under “*caTools*” package of the R statistical software. The filtering consists in eliminating the interference related to the experimental conditions and the electronic noise of the measuring instrument.

Spectral reflectance was measured in the wavelength range of 342 - 1060 nm. Since the UV band is of a little interest for the soil spectral study, we have restrained the spectral band to 400 - 1000 nm with a spectral resolution of 0.5 nm. **Figure 2(b)** and **Figure 2(c)** shows the effect of filtering on the spectrum of a sample chosen randomly. The filtered spectra were then converted to spectral absorbance (A) using the following relationship:

$$A = \log(1/R) \quad (1)$$

where R is the actual spectral Reflectance.

2.6. Chemometric Analysis of the Soil Chemistry and Spectral Data

The first part of the analysis is the calibration which consists of developing a mathematical model to determine the chemical properties Y (unknown concentration) from the available spectral measurements X (spectral absorbance). The model is setup using variable X and Y information of the calibration sample set. Once the model is established, it can be used to estimate the chemical properties of unknown samples. Two regressions models have been involved in this analysis: the principal component regression (PCR) and the partial least square regression (PLSR).

Validation is the second step of the analysis which consists in evaluating the performance of the calibrated model by comparing its estimates with the reference values. During this phase, the accuracy of the model is evaluated on a set of independent samples meaning samples that did not participate in the calibration process. Thus, a good prediction implies a good quality of the calibration model. If the model appears satisfactory, it can be applied in routine analysis to analyze unknown samples.

2.6.1. PCR Model

Principal component regression (PCR) is a two-step multivariate analysis method. The first step consists of performing a principal component analysis (PCA) of the explanatory data matrix X to convert them into new data matrix: the matrix T (X-scores) and a matrix P' (X-loadings). PCA creates new orthogonal variables T (latent variables) that are linear combinations of the original X variables with the coefficients a_i .

$$X = TP' \text{ and } T = Xa_i \quad (2)$$

In the second step, a multiple linear regression (MLR) is established between the scores obtained and the measure (known) variable Y .

The Principal Component Analysis is a way of dealing with the problem of poorly conditioned matrices. The objective is to obtain a certain number of components capturing the maximum variation relative to the variables of the

matrix X while assuring the model a certain quality of prediction. PCR can be considered as a linear regression method in which the response variable is regressed on new components.

2.6.2. PLS Model

The partial least squares regression (PLSR) is based on principal components on both the independent variable X , and the dependent variable Y . The PLS model links an unknown variable Y to a block of explanatory variables X through latent variables that are linear combinations of the initial explanatory variables [13]. The latent variables explain as much as possible the covariance between both variables X and Y . The approach is to calculate the principal scores of X and Y and to set up a regression model between the scores.

The equation system below (Equation (3)) highlights how the independent data matrix X can be decomposed into a matrix T (X-score), and a matrix, P' (X-loading), plus an error matrix, E . Similarly, the dependent data matrix Y is decomposed into the matrix U (Y-scores) and a matrix Q' (Y-loadings) plus the error term F . These decompositions are made so as to maximize the covariance between the scores matrix T and U .

$$X = TP' + E \text{ and } Y = UQ' + F \quad (3)$$

The X-scores (T_i) are orthogonal and they are estimated as linear combinations of the original variables X_r . Thus, the matrix of latent variables T is a linear transformation of X .

2.7. Statistics Assessing Model Performance

Both multivariate models were validated with an independent data set representing about 1/3 of the total sample. The models performances were assessed by using some standard statistics: coefficient of determination (R^2), the Bias and the root mean square error (RMSE). These statistical measures were computed using the following formulations:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \text{BIAS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i), \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

With \hat{y}_i the values of the measurements, y_i the predicted values and \bar{y} the average of the measurements.

3. Results and Discussion

Descriptive statistics of three soil properties C, P and K are summarized in **Table 1**. The difference between the minimum and maximum values of the concentrations for these elements demonstrates their strong variability in across the sampling sites. Also statistics for the validation sample set were within the range of the statistics for calibration sample set for all the three soil properties (**Table 1**).

3.1. Performance of Models Cross-Validation

By checking at all the statistical criteria R^2 , Bias and RMSE, both models show

Table 1. Descriptive statistics of the soil reference data constituting of 587 soil samples of calibration and 290 soil samples of validation.

Soil Properties	Calibration Set				Validation Set			
	Minimum	Maximum	Mean	STD*	Minimum	Maximum	Mean	STD*
C (%/100 g)	0.01	2.63	0.53	0.34	0.01	1.96	0.47	0.31
P (ppm)	0.21	15.90	1.34	1.40	0.31	9.16	1.32	1.40
K (meq/100 g)	0.02	1.87	0.25	0.20	0.05	1.39	0.25	0.20

*STD: Standard deviation.

good calibration qualities. They have strong coefficients of determination and weak bias for all elements. The PCR gives the best calibration quality with R^2 stronger than 0.80 for all elements (Table 2). This performance is better than that found by [14] for the carbon, with $R^2 = 0.68$ in the NIR region (700 - 2498). The PLSR model also has good calibration quality with coefficients of 0.801 for potassium, 0.872 for carbon and 0.881 for phosphorus (Table 2).

However, the coefficient of determination is not the only parameter to be considered for assessing the performance of a model. The root mean square error and the Bias between the measured and predicted values are also used as statistics to evaluate the robustness of a model.

The RMSE obtained from the PCR method is 0.22, 1.00 and 0.17 respectively for carbon, phosphorus and potassium. And for the PLSR method, the RMSE is 0.21 for the C, 1 for the P and 0.15 for the K. The bias obtained from both models (0.0004 for C; 0.0017 for P and 0.0003 for K) are relatively weaker compared to their respective average values (0.51, 1.33 and 0.25).

These results can be further improved by proceeding to homogeneous distribution of the samples from different sampling sites, as some sampling sites are represented by nearly 60 samples while others sites had only 4 samples. It has been argued that the calibration of predictive models on a limited number of samples or on fairly homogeneous samples may limit the scope of the calibration model [15].

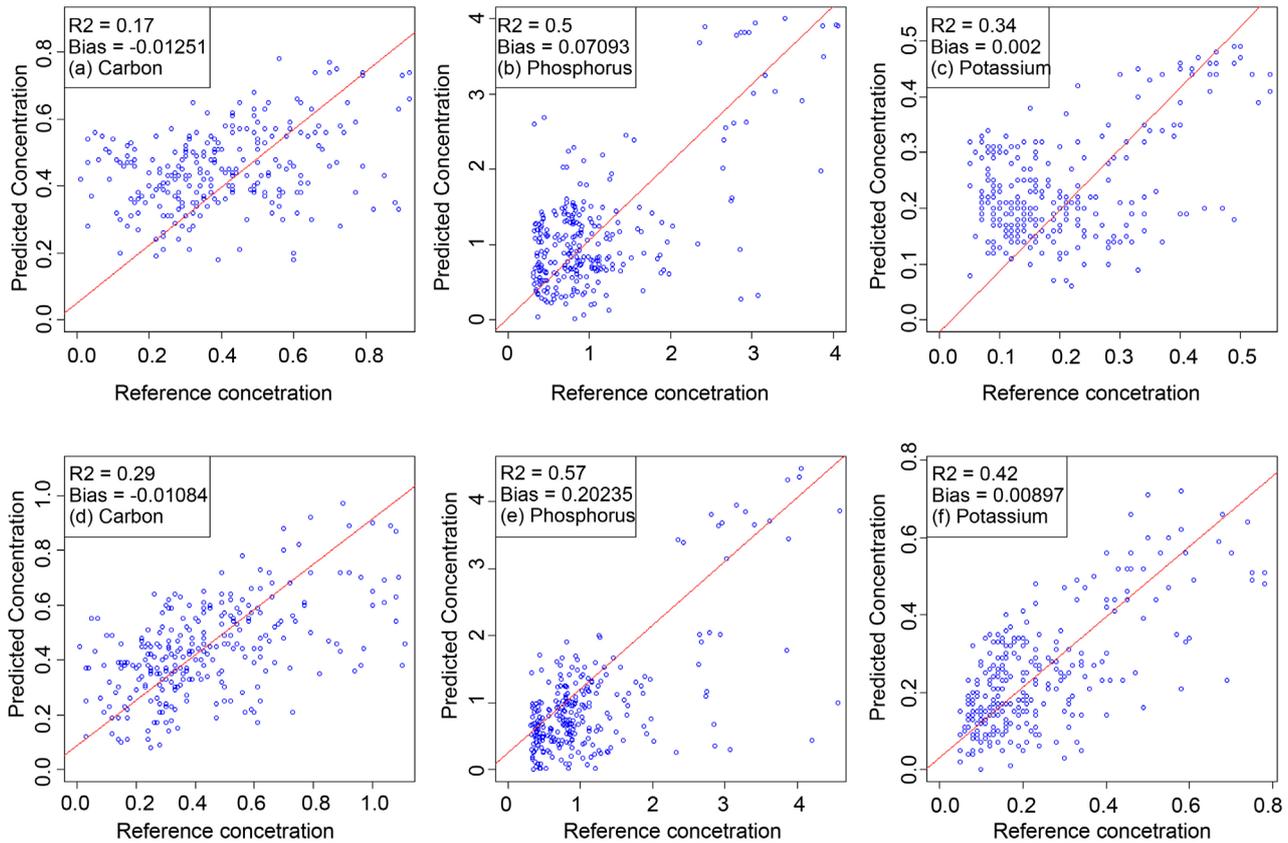
3.2. Independent Validation

It can be seen that the performance of the prediction models varies from one chemical property to another and also from one model to another. An independent validation of both multivariate models calibrated for the three chemical properties (C, K and P) reveals lower performance of prediction with regards to the cross-validation performance. The PCR model has coefficients of determination of 0.17 for carbon, 0.34 for potassium, and 0.50 for phosphorus (Figures 3(a)-(c)). These values are comparatively lesser than 0.87 (C) and 0.64 (K) obtained respectively by [16] with the band 1100 - 2500 nm and [17] with the band 400 - 2498 nm.

The independent validation of the PLSR method yielded $R^2 = 0.29$ for C, $R^2 = 0.42$ for K and $R^2 = 0.57$ for P (Figure 3(d), Figure 3(e), Figure 3(f)). Comparing

Table 2. Statistics showing the performances of the PCR and PLSR models for cross-validation (calibration) and for independent validation. All the coefficients of determination exceed the 95% of level of significant.

Soils properties	PCR						PLSR					
	Calibration			Validation			Calibration			Validation		
	R^2	Bias	RMSE	R^2	Bias	RMSE	R^2	Bias	RMSE	R^2	Bias	RMSE
C (%/100 g)	0.882	0.0004	0.22	0.17	0.0125	0.23	0.872	0.0004	0.21	0.29	0.0108	0.22
P (ppm)	0.884	0.0017	1	0.50	0.0709	1	0.881	0.0017	1	0.57	0.2023	0.95
K (meq/100 g)	0.877	0.0003	0.17	0.34	0.002	0.17	0.801	0.0003	0.15	0.42	0.009	0.16

**Figure 3.** Scatter plot with regression line showing the statistics comparing the reference values of C, P and K with their respective estimations from PCR model ((a)-(c)) and PLSR model ((d)-(f)).

to some previous finding, these performances are lower compared to $R^2 = 0.66$ (C) and $R^2 = 0.61$ (K) obtained respectively by Sorensen and [18] with the band 408 - 2492 nm and [19] with the band 400 - 2498 nm. This low performance can be attributed to the wavelength band used and also to the large extension of the study area. Indeed, the predictions of the soil components by spectroscopy may fail if the samples are collected in a very large geographical area or from different morpho-pedological contexts. Some previous findings [20] explain the failure by the difference between the original soil parent materials. However, for the potassium, our result is better than $R^2 = 0.40$ obtained by [19] in the same conditions.

Although, for some chemical, the estimates gave low performances for both

models, but it has to be noted that the PLSR assures qualities of prediction of the phenomenon better than the PCR. This is due to the fact that the PLSR components capture the information carried by the explanatory variables while paying attention to the link between the two variables.

The RMSE and the bias found are very low compared to the average of the reference data. For PCR, the RMSE obtained was 0.23 for C; 1 for P and 0.17 for K and the bias was 0.0125 for total carbon, 0.0709 for phosphorus and 0.002 for potassium. The PLSR model shows RMSE values of 0.22, 0.95 and 0.16, respectively for C, P and K; the respective biases found were -0.0108 ; 0.2023 and 0.009.

5. Conclusion

This study documents the potentiality of the VIS-NIR diffuse reflectance spectroscopy in soil study. This method of analysis is a very promising tool for soils study: rapidity, ease of measurement, without the use of chemicals and even *in situ* measurements in the field can be envisaged. Results show that the PLSR estimation over-performs the prediction of the PCR model. The independent validation reveals that VIS-NIR spectroscopy over 400 - 1000 nm has limited performance for estimating some soil properties. The prospect of using the entire spectral band of the VIS-NIR (400 - 2500 nm) for the analysis of soil properties may be considered. The creation of a spectral database by selected zone to limit the study area can be a promising solution to achieve good results. For instance, spectral soils database can be realized for specific key areas for Malian agriculture, such as "Office du Niger" and "zone CMDT" dedicated respectively for rice and cotton cultivation. Faster and easier prediction of the soil properties of these areas can be very promising. This will contribute to the development of agriculture in these areas which constitutes the major agricultural basins of Mali and have a considerable impact on the country economy.

Acknowledgements

We acknowledge the International Science Program (ISP/IPPS) for supporting the Laboratory of Optics, spectroscopy and Atmospheric Science (LOSSA) of the "Faculté des Sciences et Techniques de Bamako". Our gratitude goes to the "Laboratoire Sol-Eau-Plante de l'IER" for providing soils samples and reference chemical properties.

References

- [1] Bouman, B.A.M. (2007) Accuracy of Estimating the Leaf Area Index from Vegetation Indices Derived from Reflectance Characteristics, a Simulation Study. *International Journal of Remote Sensing*, **13**, 3069-3084. <https://doi.org/10.1080/01431169208904103>
- [2] Ziadi, N., Tremblay, G., Belanger, G., Abdi, D., St. Luce, M. and Parent, L.É. (2013) La spectroscopie dans le proche infrarouge au service de la science du sol. 27e congrès annuel de l'AQSSS "Les sols à bout de souffle?" aqsss. programme-scientifique, 76.

- [3] Sudduth, K.A. and Hummel, J.W. (1991) Evaluation of Reflectance Methods for Soil Organic Matter Sensing. *Transactions of the ASAE*, **34**, 1900-1909. <https://doi.org/10.13031/2013.31816>
- [4] Ben-Dor, E. and Banin, A. (1995) Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. *Soil Science Society of America*, **59**, 364-372. <https://doi.org/10.2136/sssaj1995.03615995005900020014x>
- [5] Genot, V., Bock, L., Dardenne, P. and Colinet, G. (2014) L'intérêt de la spectroscopie proche infrarouge en analyse de terre (synthèse bibliographique). *Biotechnology, Agronomy, Society and Environment*, **18**, 247-261.
- [6] Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J. and 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. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- [7] Udelhoven, T., Emmerling, C. and Jarmer, T. (2003) Quantitative Analysis of Soil Chemical Properties with Diffuse Reflectance Spectrometry and Partial Least-Square Regression: A Feasibility Study. *Plant and Soil*, **250**, 319-329. <https://doi.org/10.1023/A:1023008322682>
- [8] Van Groenigen, G., Muters, C., Horwath, W. and Van Kessel, C. (2003) NIR and DRIFT-MIR Spectrometry of Soils for Predicting Soil and Crop Parameters in a Flooded Field. *Plant and Soil*, **250**, 155-165. <https://doi.org/10.1023/A:1022893520315>
- [9] Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M. and Wetterlind, J. (2010) Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy*, **107**, 163-215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7)
- [10] Nduwamungu, C., Ziadi, N., Parent, L.E., Tremblay, G.F. and Thuries, L. (2009a) Opportunities for, and Limitations of, Near Infrared Reflectance Spectroscopy Applications in Soil Analysis: A Review. *Canadian Journal of Soil Science*, **89**, 531-541. <https://doi.org/10.4141/CJSS08076>
- [11] Abdi, D., Tremblay, G.F., Ziadi, N., Bélanger, G. and Parent, L.E. (2012) Predicting Soil Phosphorus-Related Properties Using Near-Infrared Reflectance Spectroscopy. *Soil Science Society of America Journal*, **76**, 2318-2326. <https://doi.org/10.2136/sssaj2012.0155>
- [12] Waruru, B.K., Shepherd, K.D., Ndegwa, G.M., Kamoni, P.T. and Sila, A.M. (2014) Rapid Estimation of Soil Engineering Properties Using Diffuse Reflectance Near Infrared Spectroscopy. *Biosystems Engineering*, **121**, 177-185. <https://doi.org/10.1016/j.biosystemseng.2014.03.003>
- [13] Desbois, D. (1999) Introduction à la régression sur variables latentes : la procédure PLS de SAS, Le cahier des techniques de l'INRA: 41-81.
- [14] Dalal, R.C. and Henry, R.J. (1986) Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Science Society of America Journal*, **50**, 120-123. <https://doi.org/10.2136/sssaj1986.03615995005000010023x>
- [15] Islam, K., Singh, B. and Mc Bratney, A.B. (2003) Simultaneous Estimation of Several Soil Properties by Ultraviolet, Visible and Near Infrared Reflectance Spectroscopy. *Australian Journal of Soil Research*, **41**, 1101-1114. <https://doi.org/10.1071/SR02137>
- [16] Chang, C.W., Laird, D.A. and Hurburgh, C.R. (2005) Influence of Soil Moisture on Near-Infrared Reflectance Spectroscopic Measurement of Soil Properties. *Soil Science*, **170**, 244-255. <https://doi.org/10.1097/00010694-200504000-00003>

- [17] Chang, C.W., Laird, D.A., Mausbach, M.J. and Hurburgh, C.R. (2001) Near-Infrared Reflectance Spectroscopy-Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal*, **65**, 480-490. <https://doi.org/10.2136/sssaj2001.652480x>
- [18] Sørensen, L.K. and Dalsgaard, S. (2005) Determination of Clay and Other Soil Properties by Near Infrared Spectroscopy. *Soil Science Society of America Journal*, **69**, 159-167. <https://doi.org/10.2136/sssaj2005.0159>
- [19] Debaene, G., Niedźwiecki, J. and Pecio, A. (2010) Visible and Near-Infrared Spectrophotometer for Soil Analysis: Preliminary Results. *Polish Journal of Agronomy*, **3**, 3-9.
- [20] Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Scott, D.E. and Coster, D.C. (1992) High Dimensional Reflectance Analysis of Soil Organic Matter. *Soil Science Society of America Journal*, **56**, 865-872. <https://doi.org/10.2136/sssaj1992.03615995005600030031x>