Optica Imaging Congress (3D, COSI, DH, FLatOptics, IS, pcAOP) © Optica Publishing Group 2023

Computational algorithm for soil organic carbon percentage estimation through NIR spectroscopy

Pablo Gomez,[1] Hans Garcia,[1] and Henry Arguello[2]

[1] *Department of Electrical Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002*

[2] *Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002*

pablo2228330@correo.uis.edu.co, hans.garcia@saber.uis.edu.co,henarfu@uis.edu.co

Abstract: This work proposes a computational algorithm based on a spectral selection NIR range for soil organic carbon percentage estimation. The results show that the proposed algorithm outperforms traditional methods by up to 0.5 R^2.

1. Introduction

In agriculture, soil organic carbon (SOC) percentage estimation is one of the most important chemical features because influences the effectiveness of farming, and optimizes growth of crop soils [1]. Traditionally, this measurement is performed by chemical methods, such as calcination or dichromate, which identify essential molecules. However, these methods present several difficulties, including high costs, time of response up to 3 months and the transfer of the soil sample to the laboratory. Therefore, to solve these difficulties, several methods have been developed for the estimation of SOC percentage by near-infrared (NIR) reflectance spectroscopy, which allows for acquiring the spectral information in the field and employing fast processing to deliver reliable results in a short period, approximately 3 minutes [2]. For acquisition of spectral information in the NIR range, there are several methods such as the whiskbroom, which allows obtaining a spectral signature for each spatial point of the sample. From this spectral signature in the NIR, multiple computational algorithms have been developed that allow the extraction of feautres and thus infer the percentage of SOC. The methods used in the state-of-the-art (SOTA) to estimate SOC from NIR spectral information are divided into 2 main groups, traditional methods, such as linear regression, random forest (RF), and support vector machine (SVM), estimate the percentage of SOC by correlating the data [3], and modern methods, such as machine learning or deep learning [4], which infer the SOC percentage through a large amount of data and exploit essential features. However, the main difficulty of the SOTA methods is their dependence on the large amount of data, and the high complexity of the computational algorithm. Therefore, the proposed computational algorithm allows obtaining the estimation of the SOC percentage through the specific choice of the spectral regions of interest, which are preprocessed and entered into a low-complexity neural network, obtaining low inference times.

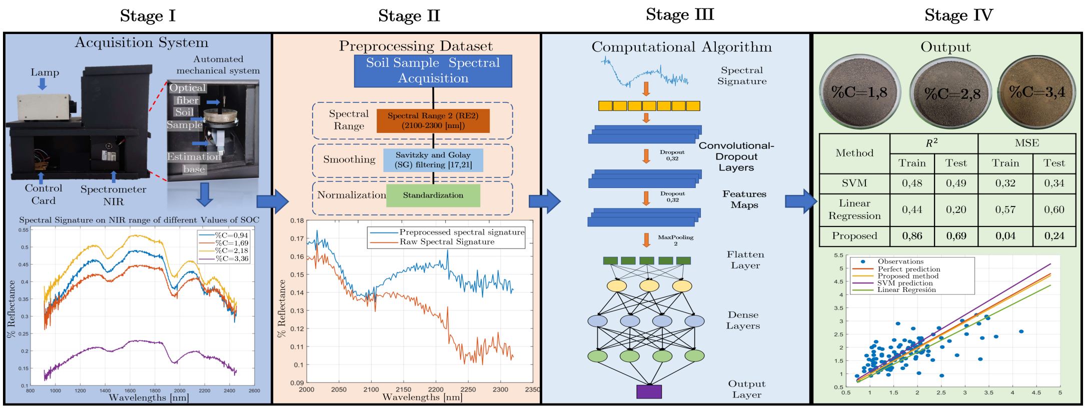


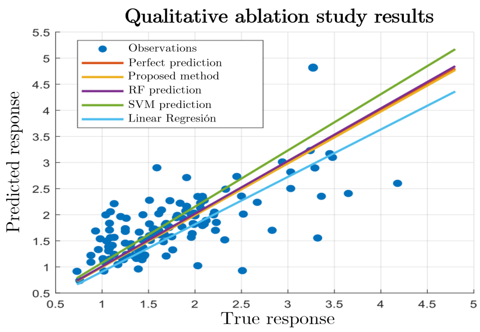
Fig. 1. Proposed method of estimation of percentage Carbon with NIR signature.

2. Proposed Method

The proposed computational algorithm is based on the principle that there are spectral regions (Red edge), in which a greater amount of information about the percentage of SOC is concentrated. In the case of the NIR range, which spans from 900-2500 [nm], it has been observed that the 2100-2300 [nm], Fig 1 stage II, range is more sensitive to this SOC chemical information. Additionally, by smoothing this spectral signature using the Savitzky-Golay method and standardizing the data, a preprocessed spectral signature is obtained that highlights the features that represent the percentage SOC. Then, this spectral signature is used as input to a low complexity 1D regression convolutional neural network which is mainly composed of 3 convolutional layers with a dropout value of 0.32, a flat layer, and a pair of dense layers, Fig.1 stage III. Concluding through this algorithm with an estimation of the percentage of SOC in different soil samples, Fig. 1 IV.

3. Simulations and Results

The proposed algorithm was tested with a dataset acquired with an optoelectronic whiskbroom system Fig. 1 stage I, which automatically acquires a total of 112 spectral signatures of each soil sample in an area of 1 inch x 1 inch, with a depth of 1 [cm]. This system allows obtaining spectral information by whiskbroom method, concluding the process with the acquired spectral signature and calculated reflectance. From this dataset, a methodology of analysis and treatment of spectral data has developed, where smoothing techniques such as Savitzky-Golay have been applied, using different function orders and window sizes for smoothing, which were [3,17,31] and [7,21,35] respectively. Additionally, 2 types of processing were tested to these spectral signatures, the first one is standardization, and the other is normalization. Thus concluding, with a wide range of possible combinations of spectral signatures preprocessing, with all the spectral range, or choosing a Red Edge that are in the range of 1300 - 1500 [nm] and 2100-2300 [nm]. Eventually, proposed method was trained for 3000 epochs for each of the above preprocessing combinations, which took a training time for each of 30 minutes, which is a small time compared to the SOTA networks for SOC estimation. Finally, we obtained a reliable SOC percentage estimate with an R^2 value of 0.69 and an MSE of 0.24, Fig. 2, due to the optimal implementation of preprocessing and Red edge selection.



Method	R_2		MSE	
	Train	Test	Train	Test
SVM	0,48	0,49	0,32	0,34
RF	0,55	0,55	0,16	0,32
Linear Regression	0,44	0,2	0,57	0,6
Proposed	**0,86**	**0,69**	**0,04**	**0,24**

Fig. 2. Results for the SOTA and proposed approach in regression for SOC percentage estimation

In addition, Table 1 shows a quantitative comparison of the results obtained by the proposed algorithm with SOTA regression methods, such as linear regression, SVM, and the RF method. Highlighting that the proposed algorithm outperforms in up to 0.5 in R^2 metric compared with the other methods to estimate the percentage of SOC. Also, a qualitative comparison of the regression is shown in Fig. 2, which shows that the regression curve obtained by the proposed method fits better the test data than the traditional methods.

4. Conclusion

A computational algorithm based on NIR selected range was proposed in this paper. Qualitative results show that from spectral range selection and preprocessing on NIR, it is possible to estimate the percentage of SOC through a low-complexity neural network, with a low training time and a reduced dataset. Remarking that the proposed approach improves traditional methods of SOTA in up to 0.5 R^2.

Acknowledgment

The authors acknowledge the Vicerrectoría de Investigación y Extension VIE at the Universidad Industrial de Santander for supporting this work under the project VIE-code 3720.

References

1. D. Beillouin and e. a. Cardinael, "A global overview of studies about land management, land-use change, and climate change effects on soil organic carbon," *Global change biology*, vol. 28, pp. 1690–1702, 2022.
2. X. Zhao and e. a. Zhao, Dongxue, "Soil organic carbon (soc) prediction in australian sugarcane fields using vis–nir spectroscopy with different model setting approaches," *Geoderma Regional*, vol. 30, p. e00566, 2022.
3. Y. Wang, S. Yang, and e. a. Yan, "Evaluation of data pre-processing and regression models for precise estimation of soil organic carbon using vis–nir spectroscopy," *Journal of Soils and Sediments*, vol. 23, pp. 634–645, 2023.
4. Q. Zhou, J. Ding, and e. a. Ge, "Estimation of soil organic matter in the ogan-kuqa river oasis, northwest china, based on visible and near-infrared spectroscopy and machine learning," *Journal of Arid Land*, vol. 15, pp. 191–204, 2023.