Science on Natural Resources and Environment
Journal homepage: tapchikhtnmt.hunre.edu.vn

INTEGRATION OF LANDSAT 8 IMAGERY AND CART MODEL FOR ESTIMATING SOIL ORGANIC CARBON IN DAK LAK PROVINCE

Duong Dang Khoi
Hanoi University of Natural Resources and Environment, Vietnam

Received 05 August 2024; Accepted 23 December 2024

Abstract

The storage potential of Soil Organic Matter (SOM) is critical for reducing CO_2 emissions. Recent advancements in remote sensing and machine learning have enabled significantly more precise prediction of SOM compared to traditional soil surveys. This study aims to examine the integration of Landsat 8 imagery and Classification And Regression Tree (CART) in estimating SOM in Dak Lak province. Landsat 8 imagery is utilized to extract spectral indices covariables that relate to SOM. The CART model was then applied to estimate SOM based on the covariables. A representative dataset of soil samples from various sites across the province was divided into training and validation subsets to evaluate the performance of the CART-based prediction. The validation result of the CART indicates that the RMSE and standard error of the model are 1.323 and 0.165, respectively. The estimation result indicates that the total amount of soil organic carbon is approximately 70.22 million tonnes of carbon in the topsoil of the province. The study provides baseline information for future estimates and carbon monitoring efforts in the topsoil of the province.

Keywords: Soil Organic Carbon (SOC); Classification And Regression Tree (CART); Landsat 8; Dak Lak.

Corresponding author, Email: ddkhoi@hunre.edu.vn

DOI: http://doi.org/10.63064/khtnmt.2024.635

1. Introduction

Soil organic carbon (SOC), also known as soil organic matter (SOM), refers to the carbon stored in the topsoil. It plays a crucial role in mitigating climate change by reducing atmospheric carbon dioxide levels [11]. However, SOM can be released into the atmosphere through natural processes like decomposition and wildfire, as well as human activities such as logging and land use change. The storage of carbon in topsoil is estimated to be about 680 billion tonnes [7], highlighting significant carbon reservoirs in soil. In addition to its role in climate regulation, SOM is vital for soil health [12, 15]. The pattern of SOM

largely varies across landscapes due to various factors, including parent material, topography, climate, vegetation, land use, and management practices [9].

Several methods are available for mapping SOM, which are categorized into field-based and remote sensing-based approaches [4, 18]. In the field-based method, soil samples are collected from various sites and analyzed in the laboratory. Geostatistical techniques are then applied to map the SOM. Although this method is considered a standard approach, it can be labor-intensive and time-consuming

The remote sensing-based method employs reflectance values obtained from satellite or airborne sensors, or derived vegetation indices, to map SOM. Reflectance in the short-wave infrared (SWIR) and near-infrared (NIR) regions has been shown to be sensitive to SOM, with reflectance decreasing as SOM increases [2]. Spectral indices derived from SWIR and NIR reflectance have been developed to detect change in SOM [14, 16]. Both the field-based and remote sensing-based methods rely on regression analysis (RA) or machine learning (ML) algorithms. The RA uses soil samples collected at specific locations to predict SOM levels at other sites. However, the accuracy of these predictions can be limited by the representativeness of the sample locations and the availability of soil samples. RA models may also be sensitive to the selection of explanatory variables used in model development. On the other hand, ML models can be a useful tool for mapping SOM. ML methods do not rely on assumptions of normal distribution of samples, allowing for the quantification of complex relationships between SOM and covariates at large scales. This makes ML models particularly suitable for estimating SOM on a per-pixel basis [4].

Among ML algorithms, tree-based algorithms have been widely used and proven to perform well in improving accuracy and decreasing uncertainty when estimating soil properties at different scales [1, 3, 10, 13, 19]. The Classification and Regression Tree (CART) model is commonly used in soil property data analysis. The CART is a robust technique because it is easy to interpret and can be used for both categorical and continuous variables. It is a non-parametric statistical approach that utilizes decision tree models to analyze and predict complex relationships between covariables and response variables. Landsat 8 provides high-resolution multispectral imagery that captures a wide range of spectral information relevant to soil characteristics. The combination of the CART model's robustness in predictive analysis and Landsat 8's detailed spectral data enables accurate and reliable SOM estimation. Additionally, the availability and cost-effectiveness of Landsat 8 data facilitate large-scale and long-term monitoring of soil organic carbon, contributing to better soil quality management.

Soil carbon sequestration brings significant benefits for provinces specializing in agricultural production because it contributes to improving soil health and supports climate change mitigation. However, quantifying soil carbon storage potential in the area has been unknown in Dak Lak province. The

aim of this study is to evaluate carbon sequestration in soils of the province. This assessment will provide valuable insights for climate change mitigation efforts. Moreover, the study will establish a baseline for the spatial distribution of soil carbon storage in forested and cultivated areas for future analysis in the province.

2. Materials and methods

2.1. Study area

Dak Lak province is situated in the Central highlands of Vietnam (Figure 1). With a total area of approximately 13,125 square kilometers, Dak Lak is one of the largest provinces in the country. The province is bordered by Gia Lai province to the North, Lam Dong province to the South, and Cambodia to the West. It is characterized by diverse topography, encompassing plateaus, mountains, and river valleys. The terrain is generally flat to gently sloping, with elevations ranging from 400 to 700 meters above sea level. The climate in Dak Lak is characterized by distinct wet and dry seasons. The rainy season lasts from May to October, while the dry season extends from November to April. The annual mean temperature ranges from 23 °C to 28 °C, with humidity varying between 78 % and 85 %. The agricultural sector is a major part of Dak Lak's economy because the province is a major producer of coffee, rubber, and pepper in Vietnam.

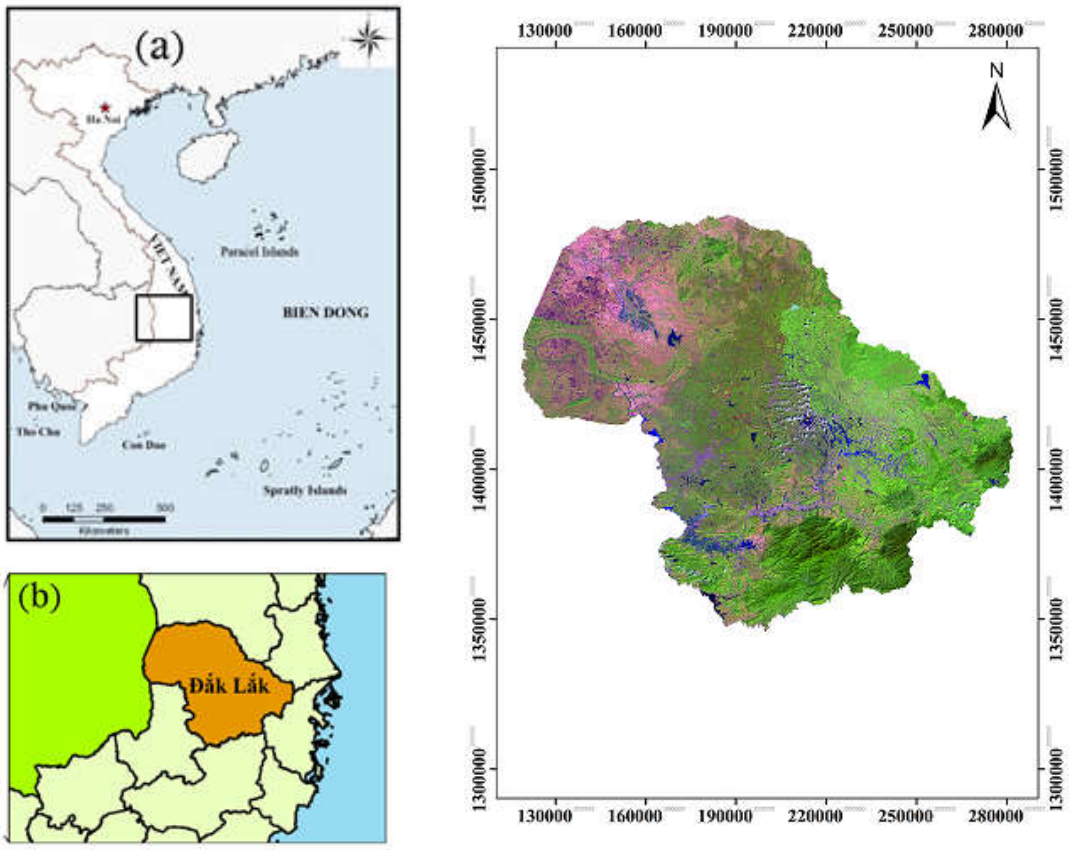


Figure 1: The site of the study area

With its extensive area, Dak Lak plays a crucial role in carbon sequestration and in soil health conservation. The province's soils possess a significant capacity for carbon storage, contributing to climate change mitigation by capturing carbon dioxide from the atmosphere and storing it in the soil. Dak Lak is known for its diverse and fertile soils, including Fluvisols, Ferralsols, and Gleysols [5, 6]. Overall, the soils of Dak Lak province hold immense importance due to their carbon storage potential and suitability for various crops, supporting both climate change mitigation efforts and the livelihoods of local communities.

2.2. Estimating soil organic matter

2.2.1. Preprocessing of Landsat 8 imagery

Applying CART involves data acquisition, preprocessing of Landsat 8 imagery, training and validation, and prediction (Figure 2). The first step is data acquisition, which involves gathering the necessary data for the model. This includes acquiring Landsat 8 imagery and obtaining an existing dataset of 677 SOM samples. The soil dataset was then divided into 478 training samples (70 %) and 199 validation samples (30 %). The Landsat 8 imagery was specifically acquired on January 20, 2020, a date typically experiencing dry weather with minimal cloud interference in the study area. The preprocessing of Landsat 8 imagery primarily involves employing the DOS (Dark Object Subtraction) technique for atmospheric correction. Atmospheric correction is an essential process in Landsat preprocessing that eliminates the atmospheric effects from the raw satellite data. To implement DOS, a dark object in the image, typically water or vegetation, which is expected to have a reflectance value of zero, is identified. The reflectance values of other objects in the image can then be adjusted accordingly based on the reflectance of the dark object.

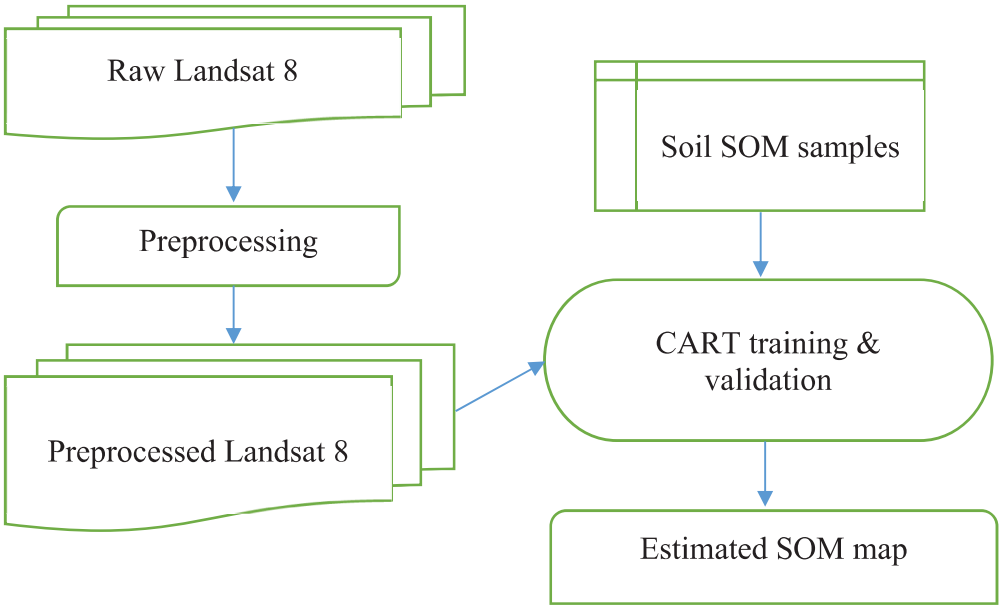


Figure 2: Flowchart of integrating Landsat 8 imagery and CART for soil organic matter estimation

2.2.2. Creation of covariables

From previous studies [2, 4, 14, 16], covariates consisting of spectral bands, NDVI, NDWI, and Clay Index have been widely chosen to estimate soil organic carbon.

NDVI is a measurement that quantifies the contrast between red and near-infrared (NIR) reflectance in vegetation. When vegetation is healthy, it tends to absorb more red light while reflecting more NIR light. The calculation of NDVI is as follows:

$$NDVI = (NIR - R)/(NIR + R) \quad (1)$$

In the Landsat 8, NDVI = (Band 5 - Band 4)/(Band 5 + Band 4).

NDWI, which stands for Normalized Difference Water Index, is a satellite-derived index that uses the near-infrared (NIR) and short-wave infrared (SWIR) bands. It compares the reflectance values of these bands to identify water bodies and assess water content in vegetation. The formula for calculating NDWI is as follows:

$$NDWI = (NIR - SWIR)/(NIR + SWIR) \quad (2)$$

In the formula [2], NIR represents the reflectance value in the near-infrared band (Band 5), and SWIR represents the reflectance value in the short-wave infrared band (Band 6). The resulting NDWI value ranges from -1 to 1, where higher values indicate a greater presence of water. Positive values suggest the presence of water, while negative values indicate the absence of water or the presence of other features such as bare soil or urban areas.

The Clay Index is determined by calculating the ratio between the Short-Wave Infrared 1 (SWIR1) and Short-Wave Infrared 2 (SWIR2) bands. It provides a measurement of the concentration of clay minerals present in the soil. In Landsat 8 imagery, the SWIR1 and SWIR2 bands are represented by bands 6 and 7, respectively.

$$Clay\ Index = SWIR1/SWIR2 \quad (3)$$

The Values to Point module in ArcGIS 10.3 was used to extract the spectral values of bands 2, 3, 4, 5, 6, 7, NDVI, NDWI, and Clay Index, along with the SOM content values of the sampled points. These variables were used as predictor variables, while the SOM content served as the dependent variable.

2.2.3. Training and validation

The next step involves training CART with a subset of the samples. In this step, the target variable (soil organic carbon) is estimated based on the values of band reflectance, NDVI, NDWI, and Clay Index. The data is divided into branches based on these predictor variable values, and decision rules are established. The model's performance is assessed using the Root Mean Square Error (RMSE) as follows:

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

where: $\hat{y}_i$ is the predicted value for the i^{th} observation in the dataset

y_i is the observed value for the i^{th} observation in the dataset

n is the sample size

2.2.4. Estimating SOM

In the final step, the calibrated CART model was used to predict the soil organic carbon content in new soil samples. This prediction was made using the spectral variables obtained from the Landsat 8 imagery. By utilizing the available data on soil bulk density for each soil unit in the topsoil and the extent of each soil unit, the total soil organic carbon stock across the entire province was estimated.

$$\text{Carbon Stock (tonnes)} = \text{SOM} \times 0.58 \times \text{A} \quad (5)$$

where SOM is soil organic matter (Mg/ha), the coefficient of 0.58 means 58 % of carbon content in SOM, and A is the area of each soil unit (ha).

3. Results

3.1. Descriptive summary of SOM content

The original dataset of samples has been divided into calibration and validation subsets. Both subsets show similar characteristics of the samples, indicating a reasonable division of the dataset. The calibration set consists of 478 samples, with an average value of 2.99 %. The minimum value in this set is 0.89 %, the maximum value is 7.47 %, and the standard deviation is 1.359. The validation set contains 199 samples, with a mean value of 2.96 %. The minimum value in this set is 0.83 %, the maximum value is 6.40 %, and the standard deviation is 1.348 %. The standard deviation reflects the spatial variability of SOM in the study area.

Table 1. Statistical description of SOM content

Set	Samples	Max (%)	Min (%)	Mean (%)	SD (%)
Whole set	677	7.47	0.83	2.98	1.355
Calibration set	478	7.47	0.89	2.99	1.359
Validation set	199	6.40	0.83	2.96	1.348

3.2. Reflectance analysis of Landsat 8 imagery

a) Calibrattion subset

Figure 3 presents the reflectance values of different bands (2, 3, 4, 5, 6, and 7) within a calibration subset derived from Landsat 8 imagery. A notable observation is the varied range of reflectance across the bands, suggesting varying degrees of electromagnetic radiation reflection from the Earth's surface.

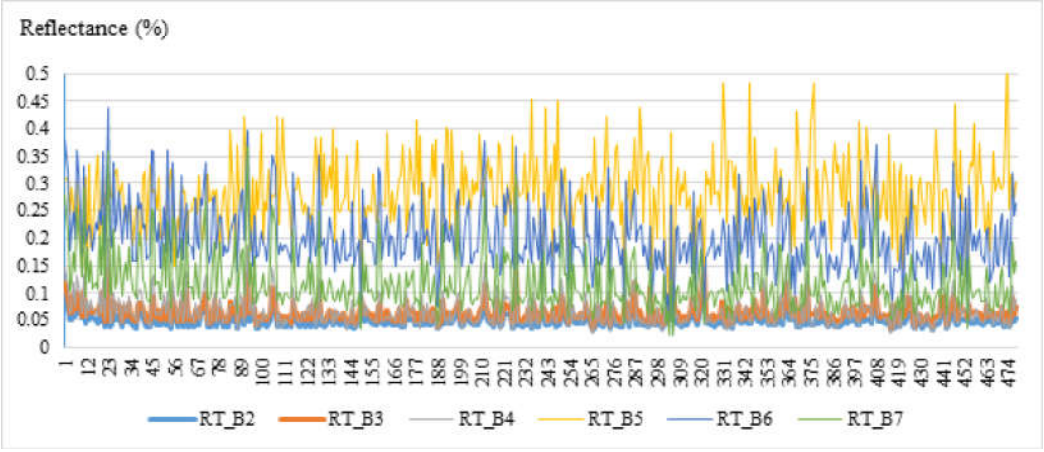


Figure 3: Reflectance of bands 2,3,4,5,6,7 of the calibration samples

Figure 4 shows the NDVI, NDWI, and Clay Index of the calibration samples. The graph illustrates the variability of the indices across the samples. The NDVI index, represented by the green line, shows a relatively stable range of values, suggesting consistent vegetation density within the calibration area. The NDWI index, depicted by the blue line, exhibits more fluctuation, indicating variation in water content. Interestingly, the Clay Index, shown in gray, presents a distinct pattern compared to the other two indices, which might be indicative of varying clay content in the soil.

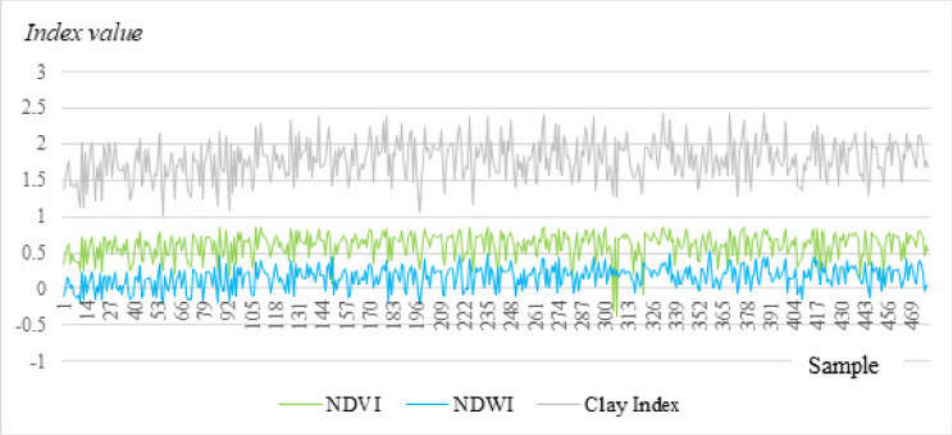


Figure 4: NDVI, NDWI and Clay Index of the calibration samples

b) Validation subset

Figure 5 presents the reflectance of a validation dataset, encompassing bands 2 to 7. The plot reveals substantial variability in reflectance across the spectral bands, suggesting heterogeneity within the validation samples. The spectral profiles of each band offer clues about the compositional properties of the materials present in the validation. The Figures 1 and 3 exhibit similar band reflectance trends. This indicates that the division of the original dataset into calibration and validation subsets is reasonable.

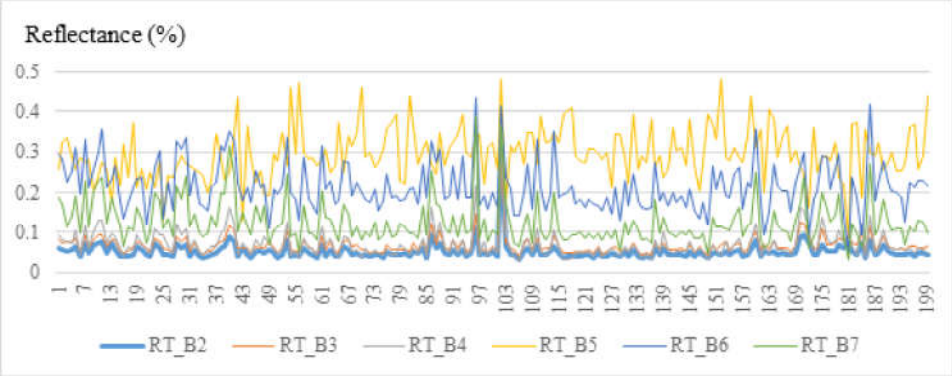


Figure 5: Reflectance of bands 2, 3, 4, 5, 6, 7 of validation samples

Figure 6 and Figure 4 indicate mostly similar trends in the NDVI, NDWI, and Clay Index of the validation samples. Over again, the graph demonstrates the variability of the indices across the samples. The NDVI index, represented by the green line, shows a quite stable range of values, suggesting consistent vegetation coverage

within the validation area. The NDWI index, depicted by the blue line, exhibits more fluctuation, indicating variations in water content. Remarkably, the Clay Index, shown in gray, presents a distinct pattern compared to the other two indices, which might be indicative of varying clay content in the soil.

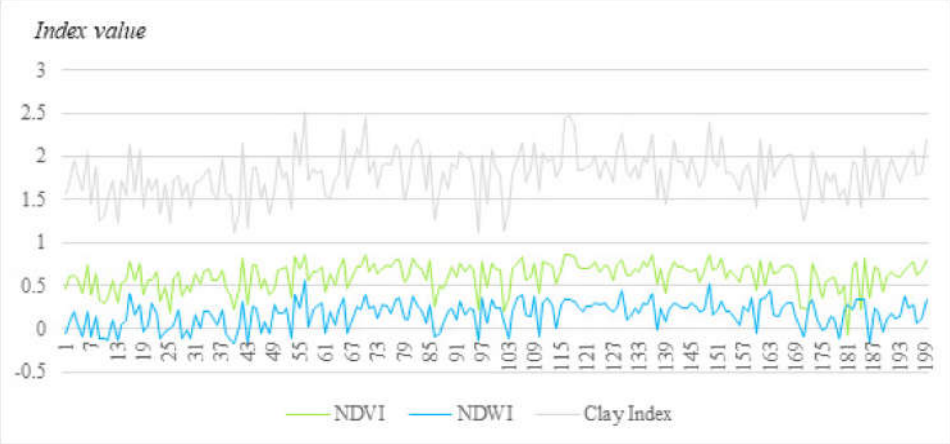
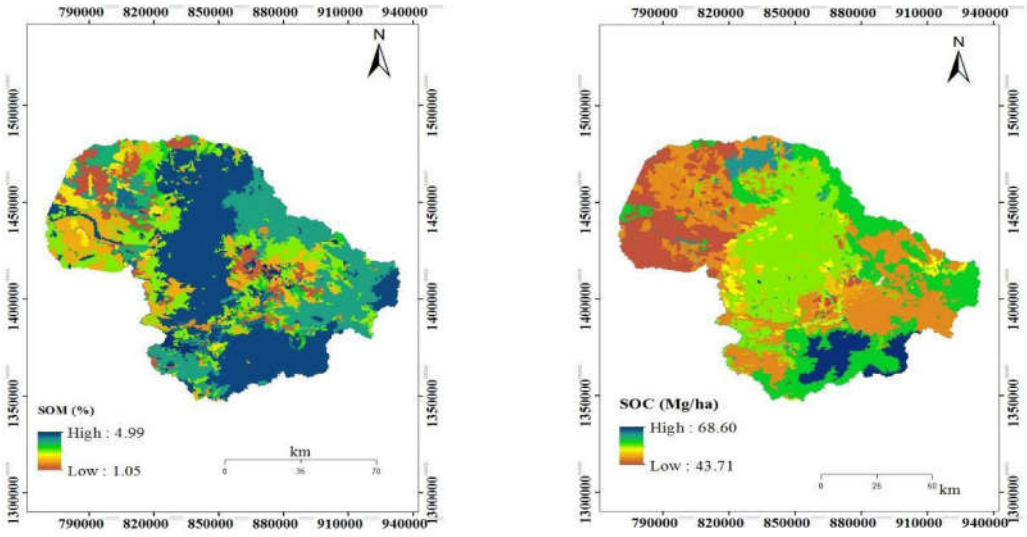


Figure 6: NDVI, NDWI and Clay Index of the validation samples

3.3. CART Validation and Prediction

To implement the training and validating of the CART algorithm, approximately 70 % of the data was used for training, while the remaining 30 % was reserved for validation purposes. The performance of the model was evaluated using RMSE. The findings reveal that the RMSE of the model during training was 1.197, with a standard error of 0.091. During validation, the RMSE and standard error were 1.323 and 0.165, respectively. These results indicate reasonably accurate performance of the model and support its application in predicting SOM in Dak Lak province using CART. Based on the validation results, the SOM map is predicted over the whole province.



(a) SOM (%) ***(b) SOM (Mg/ha)***

Figure 7: Soil organic matter estimated by CART

Soil organic carbon varies significantly in space, depending on topography, soil type, and land use in the area. Topography strongly influences SOM variation, with higher SOM content generally found in higher elevation areas where natural forests still remain, and lower SOM content on plains where coffee plantations or orchards are more common. Soil organic carbon is closely related to soil types in the province. Soils derived from organic-rich materials, such as peat and muck, tend to have higher SOM content than soils derived from igneous or sedimentary rocks, which have lower SOM content. Ferralsols are dominant in the area and contain a higher amount of SOM than other soil groups such as Fluvisols and Acrisols. The type and density of vegetation or land use covering the topsoil layer can also influence SOM content, as different crops produce varying quantities and qualities of organic residues. Forest soils tend to have higher SOM content than crop and grass soils. In general, soils under natural forests have higher SOM content than those under croplands because natural ecosystems retain more plant residues through litterfall and root exudates and have more diverse soil microbial communities that promote SOM formation and stabilization. Using the data of soil bulk density in the topsoil layer of 30 cm, together with the area extent in terms of ha of each soil, and with carbon content in OM is 58 %. The total soil organic carbon is estimated. The result indicates that the total amount of soil organic carbon is currently 70.22 millions of tonnes of carbon in the topsoil of the province.

4. Conclusions

The study findings indicate that the integration of Landsat 8 imagery and CART model is a useful tool for estimating SOM at large scale. The analysis of the covariables indicates that spectral indices such as NDVI, NDWI, Clay Index were significant for SOM estimation. The validation result of the CART indicates that RMSE and standard error of the model are 1.323 and 0.165, respectively. The calibrated CART model was applied to estimate the spatial patterns of SOM in Dak Lak province. The estimation result indicates that the total amount of soil organic carbon is approximately 70.22 millions of tonnes of carbon in the 30 cm topsoil of the province. The CART-based SOC estimation provides detailed insights into spatial distribution of SOC in the study area and supports soil conservation efforts as well as baseline information for future SOC estimates in the province.

REFERENCES

[1]. Allbed, A., Kumar, L., Aldakheel, Y.Y., (2014). *Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution images: Applications in a date palm dominated region*. Geoderma (230 - 231), 1 - 8.

[2]. Bartholomeus, H., Schaepman, M., Kooistra, L., Stevens, A., Hoogmoed, W., and Spaargaren, O., (2008). *Spectral reflectance based indices for soil organic carbon quantification*. Geoderma 145:28 - 36.

[3]. Bou Kheir, R., Bøcher, P.K., Greve, M.B., Greve, M.H., (2010). *The application of GIS based decision-tree models for generating the spatial distribution of hydromorphic organic landscapes in relation to digital terrain data*. Hydrol Earth Syst Sci.; 14: 847 - 857.

[4]. Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., et al., (2022). *Digital mapping of GlobalSoilMap soil properties at a broad scale: A review*. Geoderma 409, 115567.

[5]. Dak Lak Department of Natural Resource and Environment (2005). *Comprehensive Report of Dak Lak soil map development* (in Vietnamese).

[6]. Dak Lak Department of Natural Resource and Environment (2019). *Comprehensive Report of Land Degradation Assessment in the year 2019* (in Vietnamese).

[7]. FAO (2018). *Global soil organic carbon map: Technical report*. FAO, Rome.

[8]. IPCC (2006). *IPCC Guidelines for National Greenhouse Gas Inventories*. Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T., Tanabe K., (eds). Published: IGES, Japan.

[9]. Jenny, H., (1994). *Factors of Soil Formation: A System of Quantitative Pedology*. Courier Corporation: North Chelmsford, MA, USA.

[10]. Lamichhane, S., Kumar, L., Wilson, B., (2019). *Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review*. Geoderma, 352, 395 - 413.

[11]. Lal, R., (2004). *Soil carbon sequestration impacts on global climate change and food security*. Science 304:1623 - 1627.

[12]. Lal, R., (2006). *Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands*. Land Degrad. Dev. 17:197 - 209.

[13]. Pastick, N.J., Rigge, M., Wylie, B.K., Jorgenson, M.T., Rose, J.R., Johnson, K.D., et al., (2014). *Distribution and landscape controls of organic layer thickness and carbon within the Alaskan Yukon River Basin*. Geoderma. Elsevier B.V.; 230 - 231: 79 - 94.

[14]. Peón, J., Recondo, C., Fernández, S.F., Calleja, J., De Miguel, E., Carretero, L., (2017). *Prediction of topsoil organic carbon using airborne and satellite hyperspectral imagery*. Remote Sens. 9:1211.

[15]. Reeves, D., (1997). *The role of soil organic carbon in maintaining soil quality in continuous cropping systems*. Soil Tillage Res. 43:131 - 167.

[16]. Rossel, R.V., Walvoort, D., McBratney, A., Janik, L.J., Skjemstad, J., (2006). *Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties*. Geoderma 131:59 - 75.

[17]. Therneau, T.M., Atkinson, E.J., (2018). *An introduction to recursive partitioning using the RPART Routines*. https://cran.rproject.org/web/packages/rpart/vignettes/longintro.pdf (Accessed on 28 May 2023).

[18]. Zhang, G., Liu, F., Song, X., (2017). *Recent progress and future prospect of digital soil mapping: A review*. J. Integr. Agric., 16, 2871 - 2885.

[19]. Zhou, Y., Xue, J., Chen, S., Zhou, Y., Liang, Z., Wang, N., Shi, Z., (2020). *Fine-resolution mapping of soil total nitrogen across China based on weighted model averaging*. Remote Sens., 12, 85.