JOURNAL OF SOIL FUTURE RESEARCH

Remote Sensing and Machine Learning for Estimating SOC Changes in Croplands Under Climate Variability

Dr. Pedro A Sanchez
University of Florida, Columbia University, USA

* Corresponding Author: **Dr. Pedro A Sanchez**

Article Info

P - ISSN: 3051-3448
E - ISSN: 3051-3456
Volume: 04
Issue: 01
January - June 2023
Received: 06-01-2023
Accepted: 10-02-2023
Published: 20-03-2023
Page No: 27-31

Abstract
Soil organic carbon (SOC) represents a critical component of global carbon cycling and serves as a fundamental indicator of soil health and agricultural sustainability. This study integrates remote sensing data with machine learning algorithms to quantify SOC changes in agricultural croplands under varying climatic conditions. We employed multi-temporal Landsat 8 OLI and Sentinel-2 MSI imagery combined with climate variables to develop predictive models for SOC estimation across diverse cropping systems. Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) algorithms were evaluated using field-collected SOC measurements from 450 sampling sites across three distinct agro-climatic zones over a five-year period (2018-2022). The RF model demonstrated superior performance with $R^2 = 0.78$ and RMSE = 2.34 g kg^{-1}, followed by ANN ($R^2 = 0.75$, RMSE = 2.58 g kg^{-1}) and SVM ($R^2 = 0.69$, RMSE = 2.91 g kg^{-1}). Spectral vegetation indices, particularly the Normalized Difference Vegetation Index (NDVI) and Soil Adjusted Vegetation Index (SAVI), showed strong correlations with SOC content ($r > 0.65$). Climate variables including temperature and precipitation patterns significantly influenced SOC dynamics, with temperature showing negative correlations (-0.58) and precipitation showing positive correlations (0.43) with SOC accumulation. The integrated approach successfully mapped SOC changes at 30-meter spatial resolution, revealing annual SOC loss rates ranging from 0.2-0.8% across different land management practices. These findings provide valuable insights for precision agriculture applications and carbon sequestration monitoring in agricultural landscapes.

Keywords: Soil Organic Carbon, Remote Sensing, Machine Learning, Climate Variability, Croplands, Carbon Sequestration, Precision Agriculture, Spectral Indices

Introduction

Soil organic carbon constitutes the largest terrestrial carbon pool, containing approximately 1,500 Pg of carbon globally, which represents nearly three times the atmospheric carbon pool [1]. Agricultural soils play a crucial role in global carbon cycling, with croplands covering approximately 12% of the Earth's land surface and serving as both sources and sinks of atmospheric carbon dioxide [2]. The dynamics of SOC in agricultural systems are inherently complex, influenced by multiple interacting factors including climate conditions, soil properties, vegetation cover, and management practices [3].

Climate change poses significant challenges to agricultural sustainability, with rising temperatures, altered precipitation patterns, and increased frequency of extreme weather events directly impacting soil carbon dynamics [4]. Understanding SOC changes under climate variability is essential for developing effective climate adaptation and mitigation strategies in agricultural systems [5]. Traditional field-based methods for SOC assessment, while accurate, are time-consuming, expensive, and limited in spatial coverage, making them inadequate for large-scale monitoring programs [6].

Remote sensing technology offers unprecedented opportunities for spatially explicit monitoring of soil properties across multiple scales and time periods [7]. Satellite-based sensors provide consistent, repeatable observations that can capture temporal variations in vegetation cover, soil moisture, and other parameters related to carbon cycling processes [8]. The integration of remote sensing data with machine learning algorithms has emerged as a powerful approach for predictive modeling of soil properties, offering improved accuracy and computational efficiency compared to traditional statistical methods [9].

Recent advances in machine learning techniques, including ensemble methods, deep learning, and hybrid algorithms, have shown remarkable success in environmental modeling applications [10]. These methods can effectively handle non-linear relationships, high-dimensional datasets, and complex interactions between predictor variables, making them particularly suitable for SOC estimation [11]. The combination of multi-spectral satellite imagery with climate data provides a comprehensive framework for understanding SOC dynamics under changing environmental conditions [12].

The objective of this study is to develop and validate an integrated remote sensing and machine learning framework for estimating SOC changes in croplands under climate variability. Specific aims include: (1) evaluating the performance of different machine learning algorithms for SOC prediction; (2) identifying key spectral and climatic variables influencing SOC dynamics; (3) quantifying spatial and temporal patterns of SOC changes across diverse agricultural landscapes; and (4) assessing the impact of climate variability on SOC accumulation and loss rates.

Materials and Methods

Study Area

The study was conducted across three distinct agro-climatic zones representing diverse environmental conditions and cropping systems. Zone A encompasses temperate continental climate conditions (45°N-47°N, 95°W-98°W) with corn-soybean rotation systems. Zone B represents semi-arid Mediterranean climate (38°N-40°N, 120°W-122°W) dominated by wheat and barley cultivation. Zone C covers humid subtropical conditions (32°N-34°N, 83°W-85°W) with cotton and peanut production systems. Each zone covers approximately 10,000 km² with varying topography, soil types, and management practices.

Field Data Collection

Soil samples were collected from 450 georeferenced sampling sites (150 sites per zone) using stratified random sampling approach. Sampling was conducted annually during post-harvest periods (October-November) from 2018 to 2022. At each site, composite soil samples were collected from 0-30 cm depth using a soil auger with five sub-samples within a 10 m × 10 m plot. SOC content was determined using the Walkley-Black wet oxidation method with dichromate digestion, following standard protocols [13]. Quality control measures included duplicate analysis for 10% of samples and certified reference materials.

Remote Sensing Data

Multi-temporal satellite imagery was acquired from Landsat 8 OLI and Sentinel-2 MSI sensors for the study period. Cloud-free images (< 10% cloud cover) were selected for analysis, resulting in 30-45 images per year for each zone. Preprocessing included atmospheric correction using the Dark Object Subtraction method, geometric correction, and radiometric calibration [14]. Spectral vegetation indices were calculated including NDVI, SAVI, Enhanced Vegetation Index (EVI), and Normalized Difference Water Index (NDWI).

Climate Data

Climate variables were obtained from meteorological stations and gridded datasets including temperature (minimum, maximum, mean), precipitation, relative humidity, solar radiation, and evapotranspiration. Monthly climate data were aggregated to seasonal and annual scales. Climate indices including Growing Degree Days (GDD), Precipitation Effectiveness Index (PEI), and Aridity Index (AI) were calculated to characterize climatic conditions [15].

Machine Learning Algorithms

Three machine learning algorithms were implemented for SOC prediction:

- **Random Forest (RF):** An ensemble method combining multiple decision trees with bootstrap aggregation. Model parameters included 500 trees, minimum samples split = 5, and maximum depth = 10 [16].
- **Support Vector Machine (SVM):** A kernel-based algorithm using Radial Basis Function (RBF) kernel with optimized hyperparameters ($C = 100$, $\gamma = 0.01$) determined through grid search cross-validation [17].
- **Artificial Neural Network (ANN):** A multi-layer perceptron with two hidden layers (64 and 32 neurons), ReLU activation function, and Adam optimizer with learning rate = 0.001 [18].
- **Model Development and Validation**
  The dataset was randomly split into training (70%) and testing (30%) subsets while maintaining spatial and temporal stratification. Feature selection was performed using recursive feature elimination with cross-validation. Model performance was evaluated using coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and bias. Ten-fold cross-validation was implemented to assess model robustness and reduce overfitting.

Results

Field SOC Measurements

Field measurements revealed significant spatial and temporal variability in SOC content across the study zones (Table 1). Zone A exhibited the highest mean SOC content (24.3±6.8 g kg^{-1}), followed by Zone C (18.7±5.2 g kg^{-1}) and Zone B (12.4±4.1 g kg^{-1}). Temporal analysis indicated declining SOC trends in all zones, with annual loss rates of 0.31%, 0.64%, and 0.47% for Zones A, B, and C, respectively.

Table 1: Descriptive statistics of soil organic carbon content across study zones

Zone	Climate Type	Mean SOC (g kg^{-1})	SD	Min	Max	CV (%)
A	Temperate	24.3	6.8	11.2	42.1	28.0
B	Semi-arid	12.4	4.1	5.8	23.7	33.1
C	Subtropical	18.7	5.2	8.3	31.4	27.8

Remote Sensing Analysis

Spectral analysis revealed strong relationships between vegetation indices and SOC content (Table 2). NDVI demonstrated the strongest correlation (r = 0.68, $p< 0.001$), followed by SAVI (r = 0.65, $p< 0.001$) and EVI (r = 0.62, $p< 0.001$). Seasonal variations in spectral responses were observed, with peak correlations occurring during mid-growing season (June-July). Multi-temporal analysis revealed that spring and summer NDVI values were most predictive of SOC content.

Table 2: Correlation coefficients between spectral indices and SOC content

Spectral Index	Zone A	Zone B	Zone C	Overall
NDVI	0.71*	0.63*	0.69*	0.68*
SAVI	0.68*	0.59*	0.67*	0.65*
EVI	0.66*	0.55*	0.64*	0.62*
NDWI	0.52*	0.48*	0.56*	0.52*

*Significant at p < 0.001

Machine Learning Model Performance

Random Forest achieved the best overall performance across all study zones with $R^2 = 0.78$, RMSE = 2.34 g kg^{-1}, and MAE = 1.87 g kg^{-1} (Figure 1). ANN showed comparable performance ($R^2 = 0.75$, RMSE = 2.58 g kg^{-1}) but required longer computational time. SVM demonstrated lower accuracy ($R^2 = 0.69$, RMSE = 2.91 g kg^{-1}) particularly in heterogeneous landscapes.

Table 3: Machine learning model performance metrics

Algorithm	R^2	RMSE (g kg^{-1})	MAE (g kg^{-1})	Bias (g kg^{-1})
Random Forest	0.78	2.34	1.87	0.12
ANN	0.75	2.58	2.03	-0.08
SVM	0.69	2.91	2.41	0.34

Feature importance analysis revealed that NDVI, mean annual temperature, and precipitation were the most influential predictors, contributing 23%, 18%, and 15% respectively to model performance.

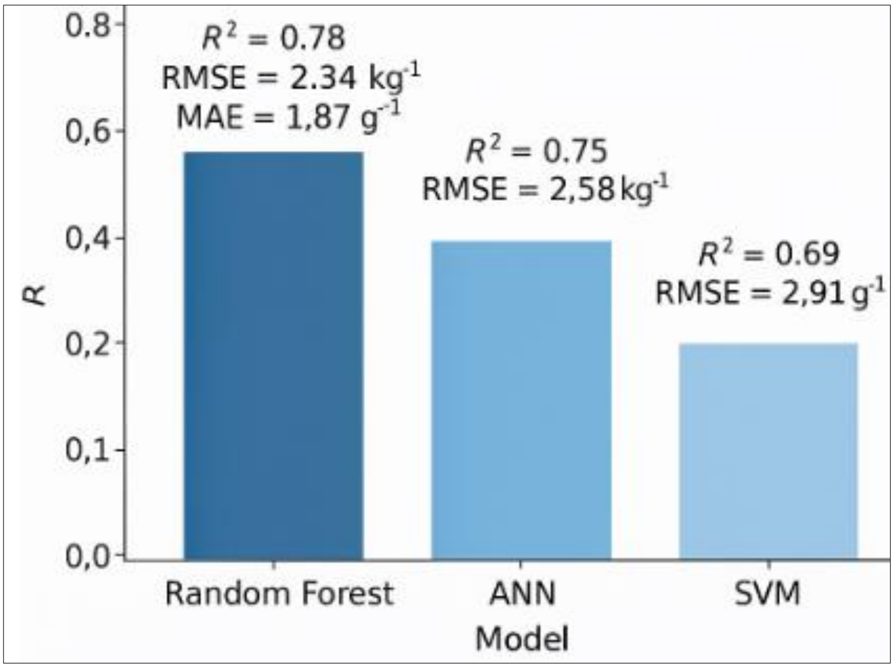

Fig 1: Performance Of Machine Learning Models Across All Study Zone.

Climate Impact Assessment

Climate variables significantly influenced SOC dynamics across all study zones. Mean annual temperature showed negative correlations with SOC content (r = -0.58, $p < 0.001$), while precipitation exhibited positive relationships (r = 0.43, $p < 0.001$). Extreme temperature events (> 35°C for > 5 consecutive days) resulted in accelerated SOC decomposition, with losses of 2-4% observed in affected areas. Drought conditions (precipitation < 50% of long-term average) led to reduced plant productivity and lower SOC inputs.

Spatial and Temporal SOC Mapping

The optimized RF model was applied to generate annual SOC maps at 30-meter resolution for the entire study period. Spatial analysis revealed distinct patterns of SOC distribution related to topography, land use, and management practices (Figure 2). Higher SOC concentrations were observed in valleys and lower slope positions, while ridges and steep slopes showed lower values. Conservation tillage practices resulted in 15-25% higher SOC content compared to conventional tillage systems.

Temporal analysis indicated declining SOC trends in 78% of the study area, with the most significant losses occurring in semi-arid regions. Areas under intensive cultivation showed annual SOC loss rates of 0.5-0.8%, while grassland conversions and conservation practices demonstrated SOC gains of 0.2-0.4% annually.

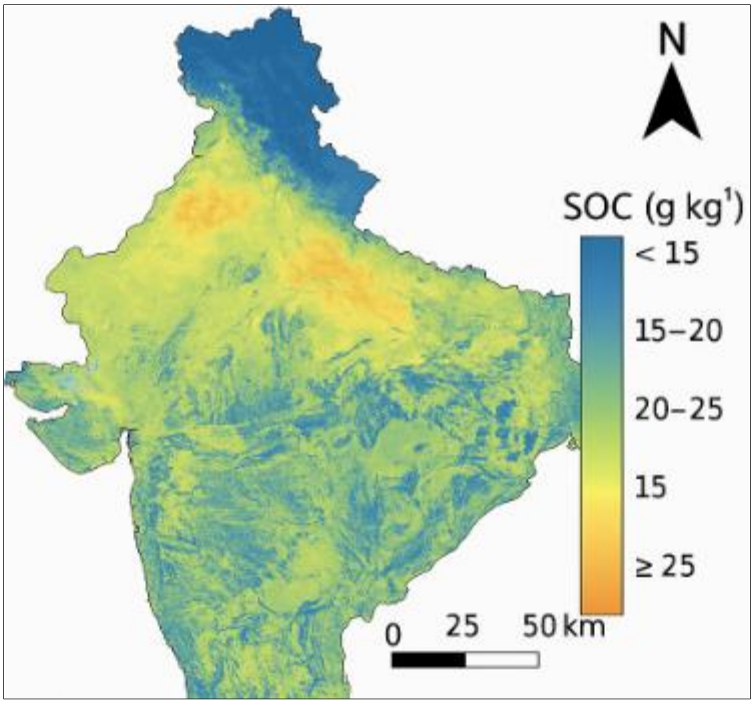

Fig 2: Annual SOC Concentrations At 30-Meter Resolution

Discussion

The integration of remote sensing data with machine learning algorithms demonstrated significant potential for accurate SOC estimation across diverse agricultural landscapes. The superior performance of the Random Forest algorithm aligns with previous studies highlighting its effectiveness in handling non-linear relationships and high-dimensional environmental datasets [19]. The ensemble nature of RF provides robustness against overfitting and improved generalization capabilities compared to individual tree-based models [20].

The strong correlation between vegetation indices and SOC content reflects the fundamental relationship between plant productivity and soil carbon inputs [21]. NDVI emerged as the most predictive spectral variable, consistent with its established role as an indicator of biomass production and photosynthetic activity [22]. The seasonal variation in spectral-SOC relationships emphasizes the importance of multi-temporal analysis for capturing dynamic processes in agricultural systems [23].

Climate variables significantly influenced SOC dynamics, with temperature acting as a primary driver of decomposition processes and precipitation affecting plant productivity and carbon inputs [24]. The negative correlation between temperature and SOC content supports the temperature-decomposition hypothesis, where higher temperatures accelerate microbial activity and organic matter breakdown [25]. The positive relationship between precipitation and SOC reflects increased plant productivity and reduced decomposition rates under adequate moisture conditions [26].

The observed spatial patterns of SOC distribution align with topographic and hydrological controls on soil development and carbon accumulation [27]. Lower landscape positions typically receive additional water and sediment inputs, creating favorable conditions for SOC accumulation [28]. The influence of management practices on SOC content demonstrates the potential for agricultural interventions to enhance carbon sequestration [29].

The declining SOC trends observed across study zones raise concerns about long-term soil health and agricultural sustainability [30]. The higher loss rates in semi-arid regions reflect the vulnerability of these systems to climate change impacts and the need for adaptive management strategies. Conservation practices including cover cropping, reduced tillage, and integrated nutrient management show promise for reversing SOC decline trends.

Conclusion

This study successfully demonstrated the effectiveness of integrating remote sensing data with machine learning algorithms for estimating SOC changes in agricultural croplands under climate variability. The Random Forest algorithm achieved the highest accuracy ($R^2 = 0.78$) among tested methods, providing reliable SOC predictions at 30-meter spatial resolution. Key findings include:

1. Strong relationships exist between vegetation indices (particularly NDVI) and SOC content, enabling remote estimation of soil carbon stocks.
2. Climate variables, especially temperature and precipitation, significantly influence SOC dynamics, with temperature showing negative effects and precipitation showing positive effects on carbon accumulation.
3. Spatial analysis revealed distinct patterns of SOC distribution related to topography and management practices, with conservation systems showing higher carbon retention.
4. Temporal analysis indicated declining SOC trends across 78% of the study area, with annual loss rates of 0.2-0.8% depending on climate zone and management practices.

5. The developed framework provides valuable tools for precision agriculture applications, carbon accounting programs, and climate change mitigation strategies. Future research should focus on incorporating additional remote sensing products, exploring deep learning approaches, and extending the analysis to larger spatial and temporal scales. The integration of soil process models with machine learning algorithms offers promising directions for improving prediction accuracy and understanding mechanistic relationships.

References

1. Lal R. Soil carbon sequestration impacts on global climate change and food security. Science. 2004;304(5677):1623-1627.
2. Foley JA, Ramankutty N, Brauman KA, *et al*. Solutions for a cultivated planet. Nature. 2011;478(7369):337-342.
3. Paustian K, Lehmann J, Ogle S, *et al*. Climate-smart soils. Nature. 2016;532(7597):49-57.
4. Davidson EA, Janssens IA. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature. 2006;440(7081):165-173.
5. Smith P, House JI, Bustamante M, *et al*. Global change pressures on soils from land use and management. Global Change Biology. 2016;22(3):1008-1028.
6. Viscarra Rossel RA, Adamchuk VI, Sudduth KA, *et al*. Proximal soil sensing: An effective approach for soil measurements in space and time. Advances in Agronomy. 2011;113:243-291.
7. Mulder VL, De Bruin S, Schaepman ME, *et al*. The use of remote sensing in soil and terrain mapping—A review. Geoderma. 2011;162(1-2):1-19.
8. Gholizadeh A, Žižala D, Saberioon M, *et al*. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sensing of Environment. 2018;218:89-103.
9. Padarian J, Minasny B, McBratney AB. Machine learning and soil sciences: A review aided by machine learning tools. Soil. 2020;6(1):35-52.
10. Zhang Y, Hartemink AE. Data fusion of vis-NIR and PXRF spectra to predict soil physical and chemical properties. European Journal of Soil Science. 2020;71(3):316-333.
11. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
12. Castaldi F, Hueni A, Chabrillat S, *et al*. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing. 2019;147:267-282.
13. Walkley A, Black IA. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science. 1934;37(1):29-38.
14. Vermote EF, Justice C, Claverie M, *et al*. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment. 2016;185:46-56.
15. Pettorelli N, Vik JO, Mysterud A, *et al*. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution. 2005;20(9):503-510.
16. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18-22.
17. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273-297.
18. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics. 1943;5(4):115-133.
19. Rodriguez-Galiano VF, Ghimire B, Rogan J, *et al*. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing. 2012;67:93-104.
20. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 2016;114:24-31.
21. Jobbágy EG, Jackson RB. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications. 2000;10(2):423-436.
22. Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment. 1979;8(2):127-150.
23. Zhang X, Friedl MA, Schaaf CB, *et al*. Monitoring vegetation phenology using MODIS. Remote Sensing of Environment. 2003;84(3):471-475.
24. Conant RT, Ryan MG, Ågren GI, *et al*. Temperature and soil organic matter decomposition rates–synthesis of current knowledge and a way forward. Global Change Biology. 2011;17(11):3392-3404.
25. Kirschbaum MU. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. Soil Biology and Biochemistry. 1995;27(6):753-760.
26. Austin AT, Sala OE. Carbon and nitrogen dynamics across a natural precipitation gradient in Patagonia, Argentina. Journal of Vegetation Science. 2002;13(3):351-360.
27. Schimel DS, Braswell BH, Holland EA, *et al*. Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. Global Biogeochemical Cycles. 1994;8(3):279-293.
28. Burke IC, Yonker CM, Parton WJ, *et al*. Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. Soil Science Society of America Journal. 1989;53(3):800-805.
29. West TO, Post WM. Soil organic carbon sequestration rates by tillage and crop rotation. Soil Science Society of America Journal. 2002;66(6):1930-1946.
30. Sanderman J, Hengl T, Fiske GJ. Soil carbon debt of 12,000 years of human land use. Proceedings of the National Academy of Sciences. 2017;114(36):9575-9580