Integrating Multi-Source Remote Sensing and Soil Attributes through Ensemble Learning for Large-Scale Soil Organic Carbon Estimation

Jayantrao Mohite[1*], Suryakant Sawant[2], Danielle Berard[3], Sonali Kulkarni[4], Ankur Pandit[5], , Dineshkumar Singh[1]

[1] Tata Consultancy Services, Mumbai, India - (jayant.mohite, dineshkumar.singh)@tcs.com
[2] Tata Consultancy Services, Pune, India - suryakant.sawant@tcs.com
[3] EMILI, Manitoba, Canada – dberard@emilicanada.com
[4] Tata Consultancy Services, Tornoto, Canada – sonali.kulkarni@tcs.com
[5] Tata Consultancy Services, Indore, India - ankur.pandit@tcs.com

Keywords: Soil Organic Carbon (SOC), Ensemble Modeling, Remote Sensing, Soil Texture, HSV Color Features, Uncertainty Analysis

Abstract

Accurate estimation of Soil Organic Carbon (SOC) is essential for sustainable land management, agricultural productivity, and climate change mitigation. This study presents a novel SOC estimation framework integrating machine learning with multispectral bands, vegetation and soil indices, topographical attributes, soil texture variables, and Hue, Saturation, and Value (HSV)-derived soil color proxies. SOC data from 180 samples collected between 2007 and 2020 across 21 agricultural fields in Manitoba, Canada, were used for model training and validation. Landsat 5, 7, and 8 data were used to derive spectral and soil indices, while SoilGrids and Shuttle Radar Topographic Mission Digital Elevation Model (SRTM-DEM) provided soil texture and topographical features. Random Forest (RF), Extreme Gradient Boosting (XGB), and a Bias Corrected–Variance Weighted (BC-VW) ensemble model were evaluated across five feature scenarios. The ensemble model achieved the best performance, with a coefficient of determination (R^2) of 0.57, Root Mean Square Error (RMSE) of 0.25, and Root Mean Square Percentage Error (RMSPE) of 7.87%, outperforming individual models. SHapley Additive exPlanations (SHAP)-based feature selection identified Clay %, Short Wave Infrared 1 (SWIR1), and Value as the most influential predictors. Independent validation using 2021 and 2023 data confirmed model robustness, with RMSPE values of 10.93% and 12.83%, respectively. The results demonstrate the importance of integrating soil-specific indices, texture, and color features with ensemble modeling for scalable and reliable large-scale SOC monitoring and carbon sequestration applications.

1. Background and Introduction

Soil Organic Carbon (SOC) is a key component of soil organic matter and an essential indicator of soil health, productivity, and resilience. It contributes to nutrient availability, water retention, and aggregation while serving as a major reservoir in the global carbon cycle, storing over 1,500 Pg of carbon in the top meter of soil, more than in the atmosphere and vegetation combined (Batjes, 1996, Lal, 2004, Stockmann et al., 2013). However, intensive land use, deforestation, and unsustainable tillage can deplete SOC and turn agricultural soils into carbon sources (Smith et al., 2008). Enhancing SOC stocks through conservation agriculture and residue management has thus become a key strategy for sustainable food systems and climate-change mitigation (Paustian et al., 2016)

Traditional SOC quantification methods such as dry combustion and wet oxidation (Nelson and Sommers, 1996, Chatterjee et al., 2009) provide high accuracy but are laborious, costly, and spatially limited. Process-based models such as RothC, CENTURY, and DNDC(Coleman and Jenkinson, 1996, Parton et al., 1994, Falloon and Smith, 2000) are designed to simulate long-term soil carbon dynamics; however, their application often requires intensive site-specific calibration and detailed climatic inputs, which restricts their scalability. As a result, remote sensing–driven digital soil mapping has gained prominence as an efficient means of capturing SOC variability over large areas (Mulder et al., 2011, Minasny et al., 2017).

Optical satellite sensors, including Landsat 8 and Sentinel-2, are extensively used due to their open accessibility and multispectral coverage spanning the visible to short-wave infrared (SWIR) region, where Soil Organic Carbon (SOC) exhibits distinct absorption characteristics(Rossel et al., 2006a, Castaldi et al., 2019). Studies have demonstrated the predictive power of vegetation and soil indices derived from these sensors, for instance, Zhang (Zhang et al., 2019) mapped SOC in China’s croplands using Landsat 8 NDVI time series, while (Bouasria et al., 2020) achieved accurate predictions in Moroccan arid regions through pansharpened Landsat imagery and machine-learning regressors. Sentinel-2 data have been equally effective, (Izurieta et al., 2022) applied Gaussian-process regression in Ecuadorian páramos, and (Kumar et al., 2018) combined Sentinel-2A bands with field data to predict SOC in the Sariska Tiger Reserve, India. Yet, the accuracy of purely optical approaches declines in cloudy or vegetation-covered regions, where bare-soil reflectance cannot be consistently captured.

To address these limitations, Synthetic Aperture Radar (SAR) observations have been used as a complementary data source. Sentinel-1’s C-band backscatter is sensitive to surface roughness and moisture properties indirectly linked to SOC. (Tripathi and Tiwari, 2022) and (Santos et al., 2023) showed that dual-polarized Sentinel-1 data improve SOC estimation, while (Yang and Guo, 2019) successfully modeled SOC and bulk density in coastal wetlands using radar signals where optical imagery was limited. However, radar–SOC relationships are indirect and require additional calibration; integrating SAR with optical data therefore offers the best compromise. Several recent studies

such as (Nguyen et al., 2022), (Zhou et al., 2022) and (Chen et al., 2024) demonstrated that data fusion of Sentinel-1 and Sentinel-2, combined with topography and climatic covariates, significantly enhances model robustness and spatial scalability. Despite these advances, multi-sensor approaches remain computationally complex, and few studies have explored principled ensemble modeling with explicit uncertainty quantification.

Another persistent limitation is the short temporal extent of ground data used in most SOC mapping studies. Many investigations rely on samples collected within one to three years ((Mohite et al., 2023a)) , which constrains models’ ability to generalize across interannual variability driven by crop rotation, residue management, or climate. Likewise, validations are often restricted to the same time frame and region, offering limited evidence of temporal or spatial transferability (Emadi et al., 2020, Mohite et al., 2023b). Addressing these issues requires longer temporal records, independent validation, and the incorporation of soil-specific features that capture intrinsic properties such as texture and color.

Soil texture, which comprises fractions of sand, silt, and clay, affects carbon stabilization through aggregate formation and mineral protection (Six et al., 2002). Soil color, conversely, acts as an optical proxy for organic matter and mineralogy. Although laboratory and field studies have linked darker soils to higher organic content (Wills et al., 2007), its application in satellite-based SOC mapping has been limited. Recent work using HSV or Munsell color spaces (Gholizadeh et al., 2020, Kome et al., 2024) suggests that color metrics derived from bare-soil composites could provide valuable complementary predictors. However, few large-scale SOC studies have systematically integrated such metrics with texture, topography, and spectral data.

In parallel, ensemble machine-learning approaches have shown strong promise in environmental prediction. Techniques such as variance-weighted averaging, Bayesian model averaging, and median voting have consistently improved SOC mapping accuracy relative to single models (Chen et al., 2020, Mishra et al., 2020, Wang et al., 2021). Nevertheless, uncertainty quantification remains underrepresented, and independent temporal validation is rarely performed. These gaps hinder the operational deployment of SOC prediction frameworks for decision-making in agriculture and carbon accounting.

To address these shortcomings, this study develops a robust and interpretable SOC estimation framework that (i) leverages multi-year ground SOC data spanning 2007–2020, (ii) integrates diverse predictors including bare-soil spectral indices, terrain attributes, SoilGrids texture, and HSV color metrics, (iii) applies tree-based learning algorithms (RF and XGB), and (iv) fuses their outputs through a Bias-Corrected Variance-Weighted (BC-VW) ensemble with explicit uncertainty evaluation. Independent temporal validation using 2021 and 2023 datasets is performed to assess generalizability. The objectives are to quantify the contribution of different feature groups, evaluate model performance and stability across scenarios, and demonstrate a scalable, uncertainty-aware approach for operational SOC monitoring over agricultural landscapes.

2. Materials and Methods

2.1 Study Area and Data Sources

The study was conducted over the EMILI Innovation Farms, a 5,500-acre commercial grain farming operation located near Grosse Isle in Manitoba, Canada. The site represents a typical agricultural landscape of the Canadian Prairies, characterized by fertile loamy soils, moderate slopes, and a continental climate with cold winters and warm summers. Twenty-one cultivated fields were selected to represent the diversity of soil types and management practices across the farm (Figure 1).

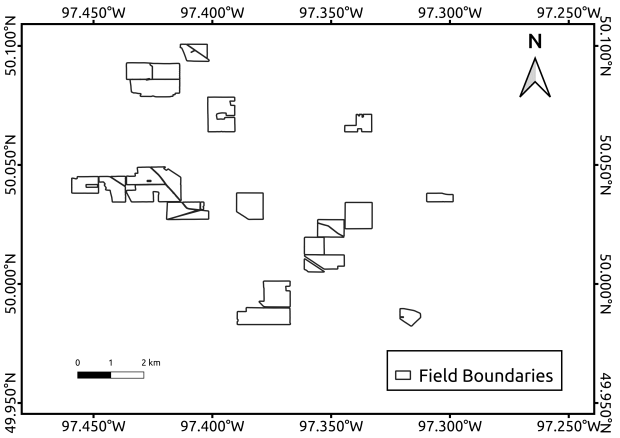

Figure 1. Study Area

A total of 180 soil samples were collected over the period 2007–2020 from the 0–15 cm soil layer, using a W-shaped composite sampling strategy to achieve representative spatial coverage within each field. Samples were air-dried, sieved to 2 mm, and analyzed for Soil Organic Carbon (SOC) content using the dry-combustion method (Nelson and Sommers, 1996). Sampling was performed after harvest (September–November) to minimize the effect of vegetation and cultivation on SOC variability.

Multispectral imagery from Landsat 5, 7, and 8 (30 m resolution) covering 2007–2020 was accessed through Google Earth Engine (GEE) (Gorelick et al., 2017). For each year, Bare Soil Composites (BaSC) were generated to isolate soil pixels under minimal vegetation and moisture influence. The procedure followed (Vaudour et al., 2021), selecting pixels with the lowest NDVI and Soil Wetness Index values during the post-harvest period. From each BaSC, raw spectral bands and sixteen derived indices (including NDVI, NDWI, SAVI, BSI, SBI, and MSI) were calculated to represent spectral and soil brightness variations.

Topographical attributes such as elevation, slope, and aspect, were derived from the SRTM DEM (30 m) (Farr et al., 2007). These variables reflect terrain driven processes affecting soil erosion and moisture redistribution. Soil texture information (sand, silt, and clay percentages) was obtained from SoilGrids (0–15 cm layer, 250 m resolution) and aggregated to the field level (Hengl et al., 2017). In addition, soil-color proxies were extracted by converting the Landsat RGB composites of BaSC imagery into HSV (Hue, Saturation, Value) color space. Hue describes dominant mineral color, Saturation expresses color purity, and Value quantifies brightness, which inversely relates to organic matter content (Rossel et al., 2006b, Rossel and Webster, 2012).

All predictor variables were spatially averaged within field boundaries to match the footprint of the ground-sampled SOC observations.

2.2 Algorithms used in this study

Two tree-based machine-learning algorithms were applied. The Random Forest (RF) algorithm (Breiman, 1999) constructs multiple decision trees using bootstrapped samples and aggregates their mean predictions, thereby reducing overfitting and handling nonlinear relationships effectively. The Extreme Gradient Boosting (XGB) algorithm (Chen and Guestrin, 2016) constructs decision trees in a sequential manner, where each new tree focuses on correcting the residual errors of the preceding ones, while regularization mechanisms are applied to limit model complexity. Both algorithms are well established for environmental and digital-soil-mapping applications.

Predictions from RF and XGB were combined using a Bias-Corrected Variance-Weighted (BC-VW) ensemble approach (Eq. 1. Each model's contribution was weighted inversely to its prediction variance, ensuring that models with higher confidence had greater influence:

$$\hat{y}_{ens} = \sum_{i=1}^{2} w_i \, \hat{y}_i, \qquad w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{2} 1/\sigma_j^2} \tag{1}$$

where $\hat{y}_i$ is the prediction and σ_i^2 the bootstrap-estimated variance of model i (RF or XGB). Models with lower variance (higher confidence) receive larger weights in the ensemble average.

This weighting scheme reduced variance and improved overall reliability without the need for a meta-learner.

2.3 Modeling Framework

In this study, a novel methodological framework was developed (Figure 2) to estimate Soil Organic Carbon (SOC) by integrating ground-based observations from 21 agricultural fields in Manitoba, Canada, with machine-learning models and multi-source predictor datasets. The dataset consisted of 180 SOC samples collected between 2007 and 2020, covering a diverse temporal span. The analysis framework integrated raw bands and spectral indices derived from Landsat 5, 7, 8 satellites, topographical features, soil texture, and soil color proxies extracted from satellite data. Topographical predictors, such as elevation, slope, and aspect, were derived from SRTM DEM data at 30 m resolution and aggregated at the field level. Landsat 5, 7, and 8 imagery was obtained through Google Earth Engine (GEE), and annual bare-soil composites were generated using NDVI-based filtering combined with soil-moisture refinement to isolate high-quality soil pixels. From this composite, raw spectral bands and indices were extracted. Soil texture attributes, including sand, silt, and clay, were obtained from SoilGrids, while soil color proxies were represented using HSV bands derived from Landsat RGB data. To systematically assess the contribution of these predictors, five feature scenarios (Table 1) were tested. These included combinations of raw Landsat bands, derived indices, topographical features, soil texture, and HSV-based soil color proxies. This approach allowed a stepwise evaluation of predictor relevance and their cumulative impact on SOC estimation. For model development, two machine learning algorithms, RF and XGB, were implemented. Model hyperparameters were optimized using a three-fold cross-validation approach on 80% of the training data, ensuring robust performance while minimizing overfitting. The remaining 20% of the dataset was used for independent testing to evaluate the final model performance. To further improve prediction accuracy and reduce uncertainty, predictions from RF and XGB were combined using BC-VW ensemble averaging. This approach assigns weights to individual model outputs according to their inverse prediction variances, so that estimates associated with higher confidence have a greater influence on the final ensemble prediction. Model performance was evaluated using statistical metrics including coefficient of determination (R^2), Root Mean Square Error (RMSE), Root Mean Square Percentage Error (RMSPE), and Bias to assess the accuracy, error magnitude, and any systematic deviations in predictions during model training and testing. Finally, uncertainty analysis was conducted using bootstrap sampling with 50 iterations, where the prediction variance was calculated for RF, XGB. Moreover, uncertainty for the BC-VW ensemble model was calculated using weighted uncertainties of RF and XGB. This enabled quantification of both individual model uncertainties and the combined uncertainty of the ensemble predictions.

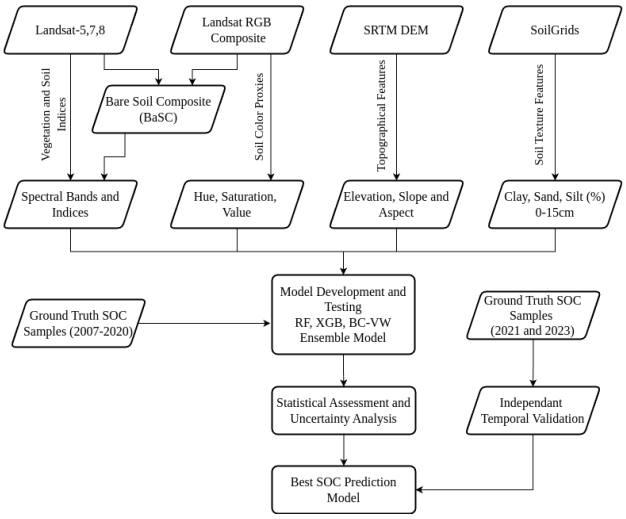

Figure 2. Overall methodology followed in this study

2.4 Independent Temporal Validation

To further evaluate the robustness and generalization ability of the developed models, we conducted independent temporal validation using SOC data collected for the years 2021 and 2023. These validation datasets were obtained for a subset of the same fields that were included during model training and testing. Independent validation was necessary to evaluate the temporal transferability of the models and to assess their predictive performance on previously unseen data. The model predictions for SOC in 2021 and 2023 were compared against the measured SOC values, and the RMSPE was calculated for both years as a quantitative performance metric. In addition to RMSPE, scatter plots were generated to visually examine the agreement between predicted and actual SOC values.

3. Results

3.1 Exploratory Analysis of Ground SOC Data

The histogram of SOC values (Figure 3) reveals a slightly right-skewed distribution, with the majority of values concentrated

Table 1. Various feature scenarios and features used in this study

SN	Feature Scenario	Bands	Count
Sc1	Landsat 5, 7, 8 bands and indices	B, G, R, NIR, SWIR1, SWIR2, TIR, NDVI, NDWI, SAVI, BSI, SBI, MSI, Bt, Wt, CI	16
Sc2	Sc1 + Topographical features	Sc1, Elevation, Slope, Aspect	19
Sc3	Sc2 + Soil Texture bands	Sc2, Clay %, Sand %, Silt %	22
Sc4	Sc3 + Soil Color proxy features	Sc3, Hue, Saturation, Value	25
Sc5	Selected top 16 features	Top selected features	–

between 3.0% and 4.0%. The highest frequency of observations occurs between 3.5% and 4.0%, indicating that the majority of samples are concentrated within this moderate SOC range. Lower frequencies are noted at the extremes, with fewer samples below 2.5% or above 4.5%, suggesting variability driven by soil characteristics, management practices, or environmental factors. Field-wise variability in SOC values is visualized through a boxplot (Figure 4), which highlights substantial differences in SOC distributions among fields. Some fields exhibit narrow ranges of SOC values, indicating homogeneity within those fields, while others display broader ranges, reflecting greater variability. For instance, fields like Field 3 and Field 7 show relatively stable SOC distributions, whereas fields like Field 15 and Field 31 exhibit wider variability, including outliers that may result from localized management practices or environmental conditions. This field-specific heterogeneity emphasizes the need for spatially adaptive modeling approaches to accurately capture SOC dynamics.

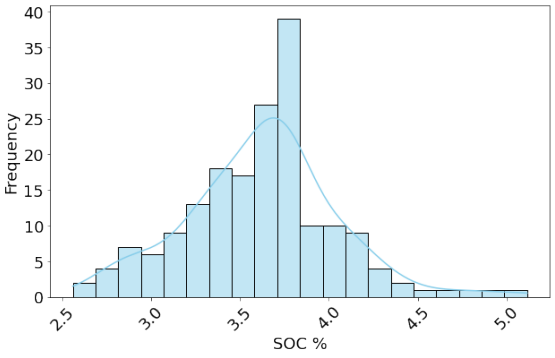

Figure 3. Histogram distribution of SOC (%)

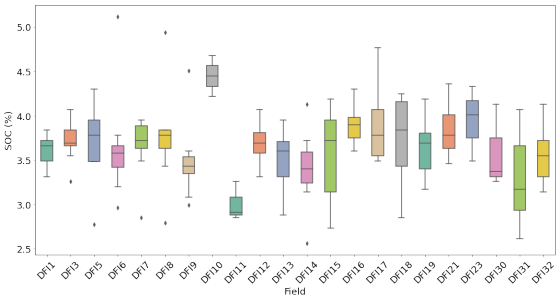

Figure 4. Field wise boxplot of SOC (%) values

3.2 Model Performance Across Feature Scenarios

3.2.1 Performance of RF The Random Forest (RF) model demonstrated varying levels of performance across the five feature scenarios, as summarized in Table 2. Scenario 1 (Sc1), which utilized all Landsat 5, 7, 8 bands and indices, yielded an R2 of 0.32, RMSE of 0.36, and RMSPE of 13.87%. This baseline scenario highlights the moderate predictive ability when only basic spectral data is used. Introducing topographical features in Scenario 2 (Sc2) slightly improved performance, reducing RMSE to 0.35 and RMSPE to 10.02%, with a stable R2 of 0.32. The addition of soil texture features in Scenario 3 (Sc3), including Clay %, Sand %, and Silt %, further enhanced the model, achieving an R2 of 0.37, RMSE of 0.35, and RMSPE of 9.44%. The most notable improvement occurred in Scenario 4 (Sc4), which combined soil texture and soil color features (Hue, Saturation, and Value) with Landsat 5, 7, 8 bands, indices, and topographical data. Sc4 achieved the best performance, with an R2 of 0.47, RMSE of 0.32, and RMSPE of 8.81%, indicating the utility of integrating diverse predictors. The improved RMSPE in Sc4 highlights RF’s capacity to predict SOC with higher precision relative to its observed mean compared to other scenarios. In Scenario 5 (Sc5), where the top 16 features were selected using SHAP-based feature importance (Figure 5), the model maintained strong performance with an R2 of 0.44, RMSE of 0.33, and RMSPE of 9.64%. This indicates that while feature selection reduces redundancy, it might omit weaker, yet relevant, features that RF leverages to capture complex nonlinear relationships. However, Sc5’s comparable performance underscores the relevance of the selected top 16 features, including Clay %, Silt %, SWIR1, SWIR2, Value, and Slope, which emerged as the most influential predictors. Bias across all RF scenarios remained negligible, underscoring the model’s balanced predictions with no systematic overestimation or underestimation.

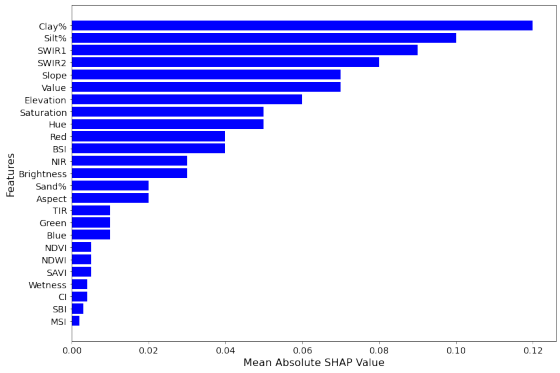

Figure 5. SHAP-based feature importance for RF

3.2.2 Performance of Extreme Gradient Boosting The XGB model exhibited superior performance compared to RF across different feature scenarios, highlighting the progressive impact of incorporating diverse features (Table 2). In Scenario 1 (Sc1), where only Landsat 5, 7, 8 bands and indices were included, the model achieved an R2 of 0.38, RMSE of 0.44, and RMSPE of 12.7%. The relatively lower performance in this scenario underscores the limited predictive power of spectral bands and in-

dices alone in SOC estimation. In Scenario 2 (Sc2), where topographical features (Elevation, Slope, Aspect) were added to the predictors, the model’s performance improved slightly, achieving an R2 of 0.40, RMSE of 0.34, and RMSPE of 9.7%. This improvement highlights the relevance of topographical features in explaining SOC variability. In Scenario 3 (Sc3), the inclusion of soil texture features (Clay %, Silt %, and Sand %) further enhanced the model’s performance, resulting in an R2 of 0.43, RMSE of 0.31, and RMSPE of 8.83%. The significant contribution of soil texture features, especially Clay % and Silt %, indicates their critical role in SOC retention and predictive accuracy. These results establish a clear trend of increasing model performance as more relevant features are incorporated, culminating in Sc4 and Sc5, where soil color proxies and SHAP-based feature selection are utilized. Scenario 5 (Sc5), with SHAP-based selection of the top 12 features, further improved performance to an R2 of 0.53, RMSE of 0.27, and RMSPE of 8.47 (Table 2). The lower RMSPE in Sc5 reflects a more precise estimation relative to the mean observed SOC values. The top 12 features selected for XGB included Clay %, SWIR1, Silt %, Value, and Elevation, which were identified as the most influential predictors for SOC (Figure 6). Bias across all scenarios for XGB was effectively zero, signifying highly unbiased predictions.

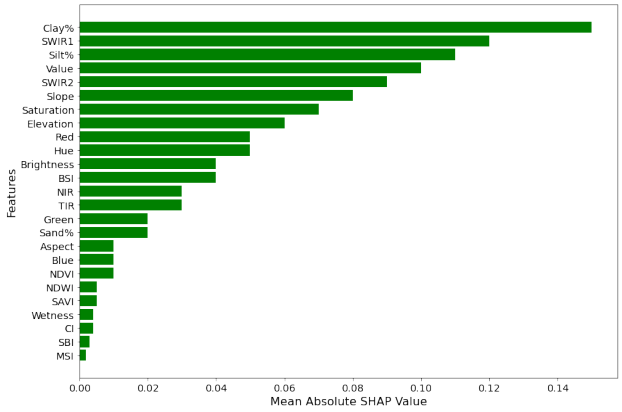

Figure 6. SHAP-based feature importance for XGB

3.2.3 Performance of Ensemble Model (BC-VW)

The ensemble model using BC-VW provided robust performance across all feature scenarios, leveraging the combined strengths of RF and XGB models (Table 2). Scenario 1 (Sc1), utilizing Landsat 5, 7, 8 bands and indices, served as a baseline, achieving an R2 of 0.38, RMSE of 0.34, and RMSPE of 9.31%. The model performance improved progressively as additional feature groups were incorporated. In Scenario 2 (Sc2), which added topographical features (elevation, slope, and aspect) to Sc1, an R2 of 0.41 and RMSE of 0.30 were recorded, with a noticeable reduction in RMSPE to 8. Scenario 3 (Sc3) incorporated soil texture components (clay, silt, and sand percentages) alongside Sc2 features, yielding an R2 of 0.46, RMSE of 0.29, and RMSPE of 8.44%. The inclusion of soil texture contributed to better predictions by capturing the relationship between SOC and soil physical properties. In Scenario 4 (Sc4), soil color proxy features (Hue, Saturation, and Value) were added to the predictors, resulting in substantial improvement with an R2 of 0.53, RMSE of 0.26, and RMSPE of 8.04%. This highlights the importance of soil color indicators as proxies for SOC. Finally, Scenario 5 (Sc5), which utilized the top features identified by SHAP analysis for RF and XGB models, provided the best results. The ensemble model in Sc5 achieved the highest R2 of 0.57, the lowest RMSE of 0.25, and RMSPE of 7.87%, outperforming RF and XGB individually. The bias remained negligible (-0.003), indicating highly accurate and unbiased predictions. Figure 7 presents the scatter plot of predicted vs. actual SOC values for best scenario from RF, XGB and ensemble model. Scenario 5 using the ensemble model is showing the closest alignment with the ideal line among all models. The results validate that the BC-VW ensemble method effectively combines the strengths of RF and XGB to enhance prediction accuracy.

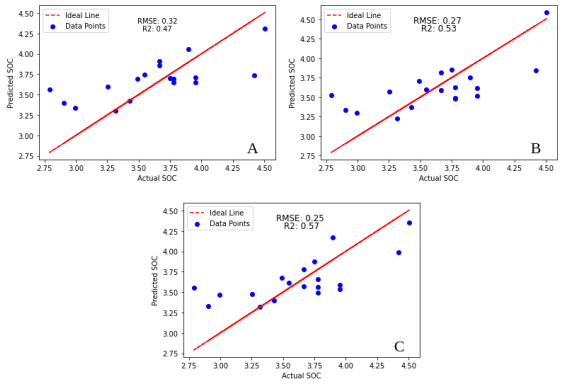

Figure 7. Scatter plot of the best performing scenario of RF - Sc4 (A), XGB - Sc5 (B) and BC-VW ensemble model - Sc5 (C)

3.2.4 Uncertainty Analysis

In this study, an uncertainty analysis was conducted to assess the reliability and robustness of predictions generated by the RF, XGB, and BC-VW-based ensemble averaging approaches. The analysis considered variability in the predictive accuracy metrics, including R2, RMSE, and RMSPE, as well as the spread of predictions through uncertainty plots. The uncertainty analysis results reveal key insights about model performance 8. For RF, the percentage uncertainty was calculated as 7.02%, with a variability of 0.14 in R2 and 0.04 in RMSE across different feature scenarios. This indicates that while RF exhibited stable performance, the inclusion of features like soil color and texture had a noticeable impact in reducing uncertainty. For XGB, the uncertainty percentage was 5.23%, demonstrating slightly lower variability compared to RF. The corresponding variability in R2 was 0.11, and RMSE variability was 0.03. These results highlight XGB’s capability to produce stable predictions with reduced uncertainty, primarily due to its strength in modeling complex interactions in the data. The ensemble model, using BC-VW-based averaging, achieved the lowest percentage uncertainty of 4.72%, with variability in R2 reduced to 0.09 and in RMSE to 0.02. These results emphasize the robustness of the ensemble model, as it combines the strengths of RF and XGB, leveraging their respective predictive capacities while mitigating their individual limitations. The reduction in uncertainty is particularly significant for SOC estimation, as it ensures more reliable predictions across spatial and temporal scales. The uncertainty plots for each model visually demonstrate the consistency of predictions (Figure 8). The scatter of predictions around the actual SOC values narrows for the ensemble model, indicating enhanced confidence in its predictions. The variability in R2 and RMSE across feature scenarios reflects the impact of including additional features, such as topographical, soil texture, and soil color features, on predictive reliability. These findings validate the effectiveness of the BC-VW-based ensemble approach in minimizing prediction uncertainty and enhancing the reliability of SOC estimation models.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-765-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 2. Model performance across algorithms and feature scenarios (RF, XGB, and Ensemble BC–VW).

SN	Scenario	Algorithm	R^2	RMSE (%)	RMSPE (%)	Bias
Sc1	Landsat 5, 7, 8 bands and indices	RF	0.32	0.36	13.87	−0.061
		XGB	0.38	0.44	12.70	0.026
		Ensemble (BC–VW)	0.38	0.34	9.31	0.035
Sc2	Sc1 + Topographical features	RF	0.32	0.35	10.02	0.008
		XGB	0.40	0.34	9.70	0.053
		Ensemble (BC–VW)	0.41	0.30	8.31	−0.005
Sc3	Sc2 + Soil Texture bands	RF	0.37	0.35	9.44	−0.048
		XGB	0.43	0.31	8.83	0.006
		Ensemble (BC–VW)	0.46	0.29	8.44	0.002
Sc4	Sc3 + Soil Color proxy features	RF	0.47	0.32	8.81	0.003
		XGB	0.50	0.31	8.79	0.003
		Ensemble (BC–VW)	0.53	0.26	8.04	−0.004
Sc5	Selected features	RF	0.44	0.33	9.64	−0.006
		XGB	0.53	0.27	8.47	−0.004
		Ensemble (BC–VW)	**0.57**	**0.25**	**7.87**	−0.003

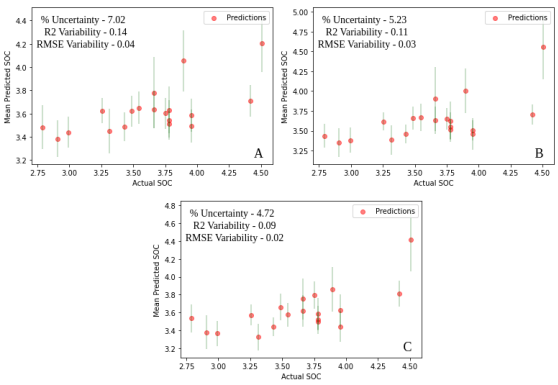

Figure 8. Uncertainty plots for various models, RF (A), XGB (B) and ensemble model (C)

3.2.5 Independent Temporal Validation Independent validation of the developed model was carried out using ground-measured SOC data from the 2021 and 2023 seasons, which were excluded from the model training and testing stages. This step is critical for evaluating the generalizability and robustness of the model when predicting SOC under completely new temporal conditions. The scatter plots for the 2021 and 2023 seasons (Figure 9) illustrate the relationship between predicted and observed SOC values, together with the corresponding RMSPE calculated for each year. For the 2021 season, the scatter plot demonstrates a reasonably strong alignment between the predicted and actual SOC values, with an RMSPE of 10.93%. This relatively low percentage error highlights the model's ability to predict SOC with reasonable accuracy, even for data temporally outside the training dataset. The distribution of points around the 1:1 line indicates that the model captures the general trend of SOC variability effectively, though some deviations are observed towards lower values of SOC. For the 2023 season, the model achieved an RMSPE of 12.83%. While this is slightly higher than the 2021 season, the performance remains satisfactory, indicating the model's capacity to maintain predictive accuracy over time. The scatter plot reveals a consistent trend in SOC prediction, with a slight increase in prediction variability for certain SOC ranges, likely due to inherent temporal and spatial variations in soil conditions.

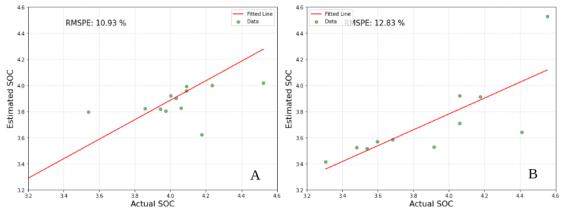

Figure 9. Scatterplot produced using ensemble model for 2021 (A) and 2023 (B) season

3.2.6 Key Findings 1. Integrating topographical, texture, and soil-color features significantly enhanced SOC prediction compared with spectral data alone.
2. The BC-VW ensemble achieved R^2 = 0.57 and RMSPE = 7.87 %, outperforming all individual models.
3. Feature-importance analysis identified Clay %, SWIR1, and Value as the most informative predictors.
4. Uncertainty reduction from 7 % (RF) to 4.7 % (ensemble) demonstrated improved confidence in predictions.
5. Independent validation confirmed temporal robustness and operational potential for large-scale SOC monitoring.

4. Discussion and Conclusions

The integration of multi-source predictors like spectral, topographical, and soil-specific variables within an ensemble learning framework substantially improved the estimation of Soil Organic Carbon (SOC). The results highlight that SOC dynamics are governed by complex soil–landscape interactions and cannot be captured accurately using spectral information alone. Combining complementary datasets and algorithms provided both higher accuracy and stronger temporal robustness.

4.1 Model Analysis and Insights

Among the individual models, Extreme Gradient Boosting (XGB) consistently outperformed Random Forest (RF), reflecting its

capacity to learn nonlinear dependencies and manage correlated predictors through gradient-based regularization. The Bias-Corrected Variance-Weighted (BC-VW) ensemble further enhanced performance by integrating the stability of RF with the sensitivity of XGB. The ensemble achieved an R^2 of 0.57 and an RMSPE of 7.87 %, while reducing prediction uncertainty to 4.72 %. These findings are consistent with previous studies where weighted or Bayesian averaging produced more reliable digital soil mapping results (Chen et al., 2020, Wang et al., 2021).

The incremental addition of predictor groups revealed the relative influence of different feature sets. Topographical factors such as elevation and slope improved model stability which could be likely due erosion and runoff processes that regulate carbon redistribution (Mishra et al., 2020, Fang et al., 2018). Texture parameters particularly Clay % and Silt % emerged as dominant variables. This could be attributed to the fact that fine particles stabilize organic matter through aggregation and physical protection. Previous studies have emphasized the importance of texture properties like Clay % and Silt % in SOC estimation due to their strong correlation with organic matter retention (Six et al., 2002, Schillaci et al., 2017). The inclusion of HSV-derived soil-color metrics introduced a novel dimension: darker soils with lower Value (V) consistently exhibited higher SOC, matching field and laboratory observations (Wills et al., 2007, Gholizadeh et al., 2020). These features likely capture subtle variations in soil brightness and chromaticity, which are indicative of organic matter content. Moreover, Short-wave-infrared bands (SWIR1, SWIR2) also ranked highly due to their sensitivity to organic-matter absorption features around 1.6–2.2 µm (Viscarra Rossel et al., 2011). Together, these predictors captured both biochemical and structural controls of SOC, explaining the ensemble’s improved performance.

The bootstrap-based uncertainty analysis confirmed the reliability of the ensemble predictions. Compared with RF and XGB individually, the ensemble produced narrower confidence intervals and lower variance across iterations, indicating greater stability. Independent validation using data from 2021 and 2023 maintained RMSPE values near 11–13 %, proving that the model learned persistent soil–landscape relationships rather than period-specific patterns. Such temporal transferability, rarely assessed in earlier SOC studies, strengthens confidence in the framework’s operational use.

4.2 Implications and Outlook

The developed approach provides a practical, scalable solution for field- to regional-scale SOC monitoring using open satellite datasets. Its accuracy and low uncertainty make it suitable for guiding soil-health assessments, precision-agriculture interventions, and carbon-sequestration reporting. Spatially explicit SOC maps derived from this framework can support site-specific nutrient management and identify low-carbon areas requiring organic-matter restoration. The implementation within cloud-based environments such as Google Earth Engine enables reproducible and cost-efficient updates, aligning with digital-agriculture and climate-action objectives.

Despite strong performance, several aspects may need improvement. The use of 30 m Landsat data may limit detection of fine-scale heterogeneity; higher-resolution or UAV imagery could enhance local detail. The exclusion of SAR features, unavoidable here due to incomplete early-year coverage, restricts information on soil structure and moisture. Future work should integrate Sentinel-1 SAR or hyperspectral observations to exploit their complementary sensitivity. Additionally, although feature selection based on SHAP improved interpretability, exploring hybrid or recursive selection schemes may balance parsimony with predictive strength. Broader validation across independent agro-ecological zones will further test spatial generalizability.

In conclusion, the proposed BC-VW ensemble framework effectively combined spectral, topographical, and soil-specific information to deliver accurate, uncertainty-aware SOC estimates. Key determinants of SOC such as Clay %, SWIR1 reflectance, and Value from HSV color space were consistently identified across models. Independent multi-year validation confirmed robustness, establishing a foundation for periodic SOC monitoring and supporting sustainable land-management, precision farming, and carbon-accounting initiatives. The study contributes a reproducible methodology that bridges remote sensing and soil science, advancing digital soil mapping under open-data and machine-learning paradigms.

References

Batjes, N. H., 1996. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 47, 151–163.

Bouasria, A. et al., 2020. Prediction of soil organic carbon using Landsat imagery and machine learning in arid lands of Morocco. *Remote Sensing*, 12(20), 3411.

Breiman, L., 1999. Random forests-random features.

Castaldi, F. et al., 2019. Evaluating the spectral sensitivity of satellite-based soil organic carbon estimation. *International Journal of Applied Earth Observation and Geoinformation*, 83, 101911.

Chatterjee, A. et al., 2009. Evaluation of soil carbon determination methods for carbon sequestration studies. *Communications in Soil Science and Plant Analysis*, 40, 1163–1178.

Chen, J. et al., 2024. Integrating Sentinel-1, Sentinel-2, and DEM data for SOC estimation using ensemble learning. *Remote Sensing*, 16(3), 501.

Chen, S. et al., 2020. Ensemble learning approaches for digital soil mapping: a review. *Soil Use and Management*, 36, 343–361.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Coleman, K., Jenkinson, D. S., 1996. RothC-26.3: A model for the turnover of carbon in soil. *Model Description and Evaluation*, 237–246.

Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., Scholten, T., 2020. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sensing*, 12(14), 2234.

Falloon, P., Smith, P., 2000. Modelling refractory soil organic matter. *Biology and Fertility of Soils*, 30, 388–398.

Fang, H., Ji, B., Deng, X., Ying, J., Zhou, G., Shi, Y., Xu, L., Tao, J., Zhou, Y., Li, C. et al., 2018. Effects of topographic factors and aboveground vegetation carbon stocks on soil organic carbon in Moso bamboo forests. *Plant and Soil*, 433(1), 363–376.

Farr, T. G. et al., 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2), RG2004.

Gholizadeh, A. et al., 2020. Soil color parameters as proxies for organic matter estimation from remote sensing. *Remote Sensing*, 12(21), 3532.

Gorelick, N. et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.

Hengl, T. et al., 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748.

Izurieta, A. A. et al., 2022. Assessment of soil organic carbon in highland ecosystems using Sentinel-2 and machine learning. *Remote Sensing*, 14(3), 565.

Kome, G. K., Enang, R. K., Yerima, B. P., Van Ranst, E., 2024. Quantitative relationships between Munsell colour attributes and organic carbon in highly weathered tropical soils. *Geoderma Regional*, 39, e00898.

Kumar, A. et al., 2018. Estimation of soil organic carbon using Sentinel-2A imagery in the Sariska Tiger Reserve, India. *Environmental Monitoring and Assessment*, 190, 680.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. *Geoderma*, 123, 1–22.

Minasny, B. et al., 2017. Soil carbon 4 per mille. *Geoderma*, 292, 59–86.

Mishra, U. et al., 2020. Evaluating ensemble machine learning approaches for mapping soil carbon at continental scales. *Environmental Modelling & Software*, 127, 104664.

Mohite, J., Sawant, S., Agrawal, R., Pandit, A., Pappula, S., 2023a. Detection of cover crop using time-series remote sensing observations. *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3430–3433.

Mohite, J., Sawant, S., Pandit, A., Pappula, S., 2023b. Estimation of soil organic carbon and bulk density for carbon stock assessment in croplands using the sentinel-1 and 2 observations. *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3494–3497.

Mulder, V. L. et al., 2011. The use of remote sensing in soil and terrain mapping—a review. *Geoderma*, 162, 1–19.

Nelson, D. W., Sommers, L. E., 1996. Total carbon, organic carbon, and organic matter. *Methods of Soil Analysis, Part 3: Chemical Methods*, Soil Science Society of America, Madison, WI, 961–1010.

Nguyen, H. et al., 2022. Synergistic use of Sentinel-1 and Sentinel-2 data for improved soil organic carbon prediction. *Remote Sensing of Environment*, 273, 112985.

Parton, W. J. et al., 1994. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochemical Cycles*, 8, 785–809.

Paustian, K. et al., 2016. Climate-smart soils. *Nature*, 532, 49–57.

Rossel, R. A. V. et al., 2006a. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131, 59–75.

Rossel, R. V., McGlynn, R., McBratney, A., 2006b. Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. *Geoderma*, 137(1-2), 70–82.

Rossel, R. V., Webster, R., 2012. Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. *European Journal of Soil Science*, 63(6), 848–860.

Santos, C. et al., 2023. Fusion of optical and SAR data for soil organic carbon estimation in tropical croplands. *Remote Sensing*, 15(1), 91.

Schillaci, C. et al., 2017. Predicting soil organic carbon using topographic and spectral covariates. *Geoderma*, 286, 12–25.

Six, J. et al., 2002. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. *Plant and Soil*, 241, 155–176.

Smith, P. et al., 2008. Greenhouse gas mitigation in agriculture. *Philosophical Transactions of the Royal Society B*, 363, 789–813.

Stockmann, U. et al., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems & Environment*, 164, 80–99.

Tripathi, R., Tiwari, R. K., 2022. Estimating soil organic carbon using Sentinel-1 SAR backscatter. *Remote Sensing Applications: Society and Environment*, 25, 100699.

Vaudour, E. et al., 2021. Using bare soil composites from Landsat time series to map soil properties. *Remote Sensing of Environment*, 254, 112259.

Viscarra Rossel, R., Chappell, A., De Caritat, P., McKenzie, N., 2011. On the soil information content of visible–near infrared reflectance spectra. *European Journal of Soil Science*, 62(3), 442–453.

Wang, J. et al., 2021. Bayesian model averaging for spatial prediction of soil organic carbon. *Geoderma*, 404, 115353.

Wills, S. A. et al., 2007. Relationship between soil color and organic carbon content in agricultural fields. *Geoderma*, 139, 322–334.

Yang, Y., Guo, X., 2019. Mapping soil properties in coastal wetlands using SAR and optical data. *International Journal of Applied Earth Observation and Geoinformation*, 81, 94–103.

Zhang, Y. et al., 2019. Mapping soil organic carbon using Landsat time series in croplands of China. *Remote Sensing*, 11(4), 406.

Zhou, T. et al., 2022. Multi-sensor data fusion for digital soil mapping of soil organic carbon. *Geoderma*, 417, 115839.