land

MDPI

Article

Soil Organic Carbon Mapping Using Multi-Frequency SAR Data and Machine Learning Algorithms

Pavan Kumar Bellam [1,2], Murali Krishna Gumma [1,*], Narayanarao Bhogapurapu [3,4] and Venkata Reddy Keesara [2]

[1] Geospatial and Big Data Sciences, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad 502324, India; b.pavan1708@gmail.com
[2] Department of Civil Engineering, National Institute of Technology, Warangal 506004, India; kvreddy@nitw.ac.in
[3] Microwave Remote Sensing Laboratory, University of Massachusetts, Amherst, MA 01003, USA; nbhogapurapu@umass.edu
[4] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA
* Correspondence: muralikrishna.gumma@icrisat.org or gummamk@gmail.com

Abstract

Soil organic carbon (SOC) is a critical component of soil health, influencing soil structure, soil water retention capacity, and nutrient cycling while playing a key role in the global carbon cycle. Accurate SOC estimation over croplands is essential for sustainable land management and climate change mitigation. This study explores a novel approach to SOC estimation using multi-frequency synthetic aperture radar (SAR) data, specifically Sentinel-1 and ALOS-2/PALSAR-2 imagery, combined with advanced machine learning techniques for cropland SOC estimation. Diverse agricultural practices, with major crop types such as rice (*Oryza sativa*), finger millet (*Eleusine coracana*), Niger (*Guizotia abyssinica*), maize (*Zea mays*), and vegetable cultivation, characterize the study region. By integrating C-band (Sentinel-1) and L-band (ALOS-2/PALSAR-2) SAR data with key polarimetric features such as the C2 matrix, entropy, and degree of polarization, this study enhances SOC estimation. These parameters help distinguish variations in soil moisture, texture, and mineral composition, reducing their confounding effects on SOC estimation. An ensemble model incorporating Random Forest (RF) and neural networks (NNs) was developed to capture the complex relationships between SAR data and SOC. The NN component effectively models complex non-linear relationships, while the RF model helps prevent overfitting. The proposed model achieved a correlation coefficient (r) of 0.64 and a root mean square error (RMSE) of 0.18, demonstrating its predictive capability. In summary, our results offer an efficient approach for enhanced SOC mapping in diverse agricultural landscapes, with ongoing work targeting challenges in data availability to facilitate large-scale SOC mapping.

Keywords: SOC; SAR; Sentinel-1; ALOS-2; machine learning; ensemble model

Received: 26 January 2025
Revised: 21 February 2025
Accepted: 23 February 2025
Published: 23 October 2025

Citation: Bellam, P.K.; Gumma, M.K.; Bhogapurapu, N.; Keesara, V.R. Soil Organic Carbon Mapping Using Multi-Frequency SAR Data and Machine Learning Algorithms. *Land* **2025**, *14*, 2105. https://doi.org/10.3390/land14112105

1. Introduction

Soil is a vital component of our planet's health, serving as the foundation for plant growth. Soil organic carbon (SOC) is a crucial indicator of soil health and fertility, playing a vital role in the global carbon cycle [1–3]. Soil's ability to sequester carbon is a key strategy for mitigating atmospheric greenhouse gases (GHGs). Soils hold an estimated 1500 to 2000 petagrams (Pg) of carbon [4–6]. This is approximately equal to the total carbon in the

atmosphere. Croplands, in particular, have significant potential for carbon sequestration ranging from 0.90 to 1.85 Pg C/yr, which is 26–53% of the target of the COP21 Paris climate summit in 2015, according to [7]. This can be achieved through improved land management practices that enhance SOC levels. According to [8], management practices such as reduced tillage, residue management, stubble retention, organic farming, crop rotation, and biochar application are effective in increasing SOC stock. These practices enhance soil quality while reducing atmospheric carbon dioxide, contributing to climate change mitigation. The spatial distribution of SOC on croplands, as well as its contribution to total carbon stock, differs substantially from that of carbon stored in above- and below-ground biomass [9].

Accurate mapping SOC in croplands, particularly during the fallow period, is essential for understanding and managing carbon dynamics [10,11]. Traditional SOC estimation methods rely on physical sampling and laboratory tests, which are labor-intensive, time-consuming, costly, and challenging to scale across large geographic areas. Additionally, these methods can be destructive to the soil. Remote sensing has emerged as a promising alternative to SOC estimation. Techniques such as vegetation indices correlate vegetation health with SOC levels, as healthier vegetation often indicates a higher SOC content [12,13]. Hyper-spectral and multi-spectral imaging capture soil reflectance across various wavelengths, enabling SOC identification through specific spectral signatures [14,15]. Optical remote sensing maps bare soil areas, where direct measurements of soil reflectance can be correlated with the SOC content, providing accurate SOC estimates [16]. Microwave remote sensing is less influenced by surface conditions such as soil moisture, roughness, and vegetation cover, which can significantly affect optical data accuracy [17]. Integrating microwave remote sensing provides significant advantages over optical methods. Additionally, microwave sensors operate in all weather conditions and at any time of day, ensuring continuous data acquisition. This capability is crucial for monitoring SOC in regions with frequent cloud cover or varying light conditions, where optical sensors often face limitations [18]. Furthermore, microwave data enable deeper soil penetration, allowing for a more comprehensive assessment of SOC distribution and dynamics. These attributes make microwave remote sensing a robust tool for SOC estimation, improving the data reliability and temporal resolution.

The advent of remote sensing technologies, particularly synthetic aperture radar (SAR) from satellites like Sentinel-1 and ALOS-2/PALSAR-2, has revolutionized our ability to monitor SOC across extensive landscapes. Polarimetric SAR data are highly sensitive to soil and vegetation roughness and moisture conditions [19]. These properties enable SAR data to provide valuable insights into SOC distribution, enhancing large-scale monitoring capabilities. By leveraging time-series data from these satellites, we can capture dynamic changes in soil properties over time. This is particularly beneficial for SOC estimation, as it allows us to account for seasonal variations and enables long-term monitoring of organic carbon dynamics. When integrated with ground-truth data, these datasets form a robust foundation for SOC modeling.

Machine learning approaches such as Support Vector Machines (SVMs), Gradient Boosting Machines (GBMs), and neural networks are valuable tools in environmental modeling [20]. These models handle large datasets and multiple input variables, effectively capturing the non-linear relationships in data derived from remote sensing technologies [21]. They can be trained on various features extracted from SAR data, including backscatter values and textural indices, to predict SOC with high precision [22]. This demonstrated that multi-temporal Sentinel-1A data, when used in RF and BRT models, enhance model performance by incorporating 15 environmental variables. These models are a cornerstone of machine learning and are extensively applied in the spatial prediction of soil properties [23]. Among various machine learning models, Random Forest (RF) has shown great

promise in environmental sciences due to its robustness and accuracy in handling complex relationships between the target variable and predictors [24]. Tree-based models, unlike linear models, excel at capturing complex non-linear relationships. RF constructs multiple decision trees during training and predicts the final outcome by aggregating the mode (for classification) or the mean (for regression) of individual tree outputs. Building upon the strengths of Random Forest, this study explores ensemble methods such as stacked models and hybrid approaches that integrate RF with other machine learning techniques, including neural networks.

This study aims to assess the effectiveness of multi-frequency and multi-temporal SAR data, combined with advanced machine learning techniques, for improving SOC estimation and mapping. Integrating these techniques enhances our understanding of soil carbon dynamics and improves cropland management for effective carbon sequestration. We hypothesize that integrating a time-series of Sentinel-1 and ALOS-2/PALSAR-2 data with an Ensemble model, which leverages the strengths of multiple machine learning models, will enhance the accuracy and reliability of SOC mapping using SAR data. The integration of machine learning with remote sensing paves the way for scalable and efficient soil resource monitoring and management. By improving carbon stock estimation with this approach, better cropping practices can be implemented, ultimately enhancing carbon sequestration in croplands and contributing to climate change mitigation strategies.

2. Study Area and Dataset

This study focuses on croplands in four districts of Odisha, India: Koraput, Nabarangpur, Rayagada, and Gajapati. These districts fall under distinct agroclimatic zones. Rayagada and Gajapati belong to the agroclimatic zone of the northeastern Ghats, characterized by a cultivated area of less than 0.1 Mha. In contrast, Koraput and Nabarangpur belong to the agroclimatic zone of the eastern Ghat mountains, and parts of Koraput extend into the southern Ghat zone. These two districts have a larger cropland area exceeding 0.2 Mha.

Rice (*Oryza sativa*) is the dominant crop across all four districts, benefiting from irrigation. Koraput has the highest crop diversity, with significant cultivation of Niger (*Guizotia abyssinica*), maize (*Zea mays*), and finger millet (*Eleusine coracana*). Nabarangpur features additional major crops such as maize (*Zea mays*) and sugarcane (*Saccharum officinarum*). In Rayagada and Gajapati, cotton (*Gossypium* spp.) and maize (*Zea mays*) are the dominant crops, with cotton being grown on newly converted agricultural lands.

The integration of Sentinel-1 C-band SAR and ALOS-2/PALSAR-2 L-band SAR data in this study provides a comprehensive assessment of SOC distribution. The C-band is effective in detecting surface soil moisture and vegetation cover, key indicators of SOC levels, while the L-band penetrates vegetation, revealing underlying soil structures essential for SOC mapping. By combining multi-temporal SAR datasets with in situ measurements, this study aims to enhance the accuracy of SOC estimation across these cropland regions.

2.1. Remote Sensing Datasets

This study employed a synergistic approach using multi-frequency SAR data to enhance SOC mapping accuracy in fallow croplands. Sentinel-1's C-band SLC data, operating at a frequency of approximately 5.405 GHz, provided imagery in VV and VH polarizations. This was complemented by ALOS-2/PALSAR-2's L-band SAR data, operating at a lower frequency of 1.215 GHz, which provided HH and HV polarizations. Sentinel-1 datasets were acquired in Interferometric Wide (IW) mode with a temporal resolution of 12 days and 20 scenes. ALOS-2/PALSAR-2 datasets were acquired in ScanSAR mode every 46 days from May to June 2020, with a total of 10 scenes acquired. The imagery footprint is shown in Figure 1. In this study, the Sentinel-1 C-band frequency was effective in detect-

ing surface soil moisture and vegetation cover, both of which are key indicators of SOC levels. The longer wavelength of the ALOS-2/PALSAR-2 L-band penetrated vegetation more effectively, revealing underlying soil structures that are essential for SOC mapping, particularly in areas with residual vegetation. To focus our analysis on croplands, we used a high-resolution cropland mask from previously published work [25], clipped to the four study districts.By integrating multi-temporal remote sensing datasets with in situ measurements, this study provides a detailed analysis of SOC variability across fallow croplands. The results highlight how integrating field observations with high-resolution satellite data improves the accuracy of SOC estimation. This approach aligns with recent advancements in remote sensing, where multi-spectral and multi-temporal data enhance SOC mapping, supporting improved agricultural land management.

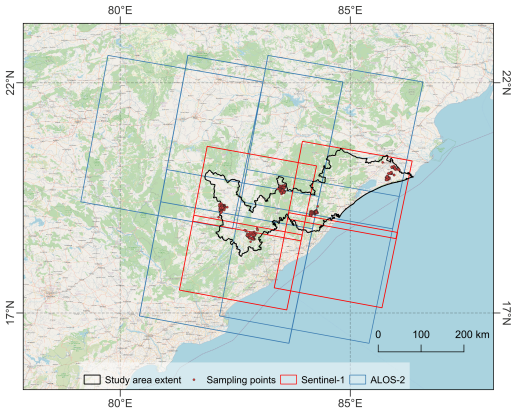

Figure 1. Map of the study area.

2.2. In Situ Datasets

The ground dataset includes SOC values (percentage) measured using the Walkley–Black method from the top soil of 15 cm depth. Initially, samples were collected from six districts in Odisha, including Puri and Khordha, resulting in over 11,500 data points. However, due to rapid urbanization as well as changes in croplands into plantations, Khordha and Puri districts were excluded to be studied separately. The final dataset comprised 8836 samples from four districts: Nawrangpur, Koraput, Rayagada, and Gajapati. These samples exhibited a narrow range of SOC values but showed a strong correlation with SLC data. The distribution of SOC sample density is illustrated in Figure 2.

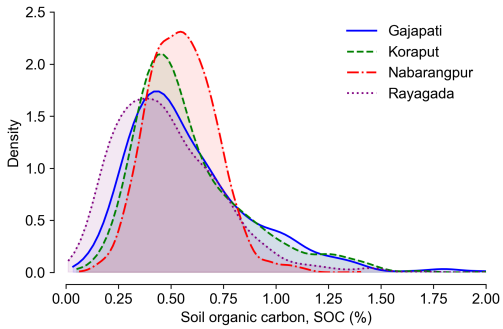

Figure 2. Density distribution of SOC across the test sites.

3. Methodology

Sentinel-1 and ALOS-2/PALSAR-2 data were acquired from the ASF and JAXA sites, respectively. SLC data were acquired in slant range geometry and contained amplitude and phase information. The preprocessing steps aimed to prepare the SLC data for further

analysis by removing noise, correcting for geometric distortions, and applying specific processing techniques for polarimetric analysis.

Preprocessing Steps:

- Radiometric Calibration: Apply radiometric calibration to convert the raw digital numbers (DNs) to radar backscatter coefficients (sigma-0) in dB.
- Thermal Noise Reduction: Remove thermal noise from the data to improve signal-to-noise ratio.
- Multilooking: Reduce speckle noise by averaging multiple looks within a resolution cell. Choose an appropriate number of looks based on the desired balance between noise reduction and spatial resolution.
- Speckle Filtering: Apply additional speckle filtering techniques (e.g., Lee filter) to further reduce noise while preserving image features.
- Terrain Correction: Apply terrain correction to account for geometric distortions caused by topography. This step is crucial for accurate analysis, especially when comparing data from different acquisition times or sensors.

In the linear dual-polarized SAR configuration, the 2×2 Hermitian positive semi-definite covariance matrix $\mathbf{C}_2$ contains the complete parametric information for a particular transmit polarization channel X and simultaneously receives X and Y, and is given as

$$\mathbf{C}_2 = \begin{bmatrix} \langle |S_{XX}|^2 \rangle & \langle S_{XX} S_{XY}^* \rangle \\ \langle S_{XY} S_{XX}^* \rangle & \langle |S_{XY}|^2 \rangle \end{bmatrix} \tag{1}$$

where $\langle \cdot \rangle$ represents the ensemble average of the covariance matrix elements, and * represents the complex conjugate.

The state of polarization of an electromagnetic (EM) wave is characterized by the degree of polarization ($0 \leq m \leq 1$). For a completely polarized EM wave, $m = 1$, and conversely, for a completely unpolarized wave, $m = 0$. In between these two extreme cases, the EM wave is treated as partially polarized, $0 < m < 1$. Barakat [26] provided an expression of m for the $N \times N$ covariance matrix. This expression is used in this study to obtain the degree of polarization m from the 2×2 covariance matrix $\mathbf{C}_2$ for dual-pol data as

$$m = \sqrt{1 - \frac{4|\mathbf{C}_2|}{(\mathrm{Tr}(\mathbf{C}_2))^2}} \tag{2}$$

where Tr is the matrix trace operator (i.e., the sum of the diagonal elements), and $|\cdot|$ is the determinant of a matrix. In addition, the two non-negative eigenvalues ($\lambda_1 \geq \lambda_2 \geq 0$) are obtained from the eigen-decomposition of the $\mathbf{C}_2$ matrix, which are then normalized with the total power Span ($\mathrm{Tr}(\mathbf{C}_2) = \lambda_1 + \lambda_2$). The eigenvalues quantify the dominance of scattering mechanisms. The eigen-decomposition of $\mathbf{C}_2$ can be expressed as

$$\mathbf{C}_2 = \mathbf{U}_2 \mathbf{\Sigma} \mathbf{U}_2^{-1}, \tag{3}$$

where $\mathbf{\Sigma}$ is a 2×2 diagonal matrix with non-negative real elements, $\lambda_1 \geq \lambda_2 \geq 0$, which are the eigenvalues of $\mathbf{C}_2$, and $\mathbf{U}_2$ is 2×2 unitary matrix, where u_i's are the unit orthogonal eigenvectors. We define the pseudo probabilities, p_i, in terms of the eigenvalues as

$$p_i = \frac{\lambda_i}{\sum_{k=1}^{2} \lambda_k}, \tag{4}$$

which we then use to derive the scattering entropy, H_{DP}, as

$$H_{\mathrm{DP}} = -\sum_{k=1}^{2} p_k \log_2(p_k). \tag{5}$$

Further, one can obtain a pseudo-scattering type parameter (θ_{DP}) [27] for dual-polarimetric SAR data as

$$\tan\theta_{\mathrm{DP}} = \frac{m\,\mathrm{Tr}(\mathbf{C}_2)\,(C_{11} - C_{22})}{C_{11}\,C_{22} + m^2\,(\mathrm{Tr}(\mathbf{C}_2))^2} \tag{6}$$

where $\theta_{\mathrm{DP}} \in [-45^\circ, 45^\circ]$ and C_{11} and C_{22} are diagonal elements of $\mathbf{C}_2$.

Model Training and Validation

A Random Forest model was initially trained using the ground data and satellite-derived features. The Random Forest algorithm, an ensemble learning method based on decision trees, excels in processing the non-linear interactions between SAR backscatter values and SOC content. By using a large number of decision trees, the Random Forest model reduces the risk of overfitting [28] and improves predictive accuracy. Each tree in the ensemble is trained on a random subset of the data, and the final prediction is made by averaging the outputs of all trees, ensuring robust and reliable SOC estimates. Studies have demonstrated that Random Forest models can effectively integrate C-band and L-band SAR data to improve SOC estimation accuracy compared to models using a single data source. This integration leverages the complementary strengths of both SAR bands, providing a more detailed and accurate representation of SOC across various landscapes and environmental conditions. Using the Random Forest model to estimate soil organic carbon (SOC) with C-band and L-band synthetic aperture radar (SAR) data has proven to be an effective approach due to the model's ability to handle large datasets and complex relationships between variables. C-band SAR data, such as those from the Sentinel-1 satellite, operate at a wavelength that is sensitive to surface roughness and vegetation structure, providing valuable information about soil conditions. L-band SAR data, on the other hand, penetrate deeper into the soil and vegetation, capturing subsurface characteristics that are crucial for accurate SOC estimation. The combination of these two bands allows for a comprehensive analysis of soil properties across different depths and surface conditions. A further Ensemble model, which was developed with the RF model, served as the foundational component in both the SNN-RF and SRF models, enabling the creation of a powerful ensemble approach by combining its robust feature learning capabilities with the strengths of neural networks and stacked ensembles for improved SOC prediction.

Ensemble Model Architecture Overview Model 1 (Sandwich Neural Network–Random Forest (SNN-RF)): This structure could be seen as a "sandwich" because you have two different types of models (NN) "sandwiching" a third model (RF) in the middle. A novel hybrid model, termed the SNN-RF, was developed. This architecture comprises three layers: The input layer receives the extracted SAR features. Hidden Layer (RF): A Random Forest (RF) layer is embedded within the neural network. The RF layer learns complex non-linear relationships between the input features and SOC, leveraging the strengths of tree-based models in capturing intricate patterns and non-linear interactions. Output Layer: A fully connected layer processes the output of the RF layer and generates the final SOC prediction. This could combine the strengths of both neural networks (for capturing complex patterns) and Random Forests (for handling feature importance and non-linearity).

Model 2 (Stacked Random Forest (SRF)): It consists of three Random Forest (RF) models, which are likely used in a stacked fashion. A traditional stacked Random Forest model was implemented as a second component of the ensemble. This model employs

multiple RF models as base learners. Each base learner independently predicts SOC based on the input features. The outputs of these individual RF models are then used as input to a meta-learner (another RF model). The meta-learner analyzes the predictions of the base learners and generates a final, more robust SOC prediction. Stacking simply combines each RF layer and its outputs in a way that minimizes the bias and variance of the predictions.

Ensemble Method: Predictions from SNN-RF and SRF models were combined for improved performance and reduced overfitting. The models were used to generate multiple predictions for each input. The final output was the average of all model predictions. Standard deviation was used to measure uncertainty and variance in predictions, indicating model confidence. The architectural flow of this ensemble model is shown in Figure 3.

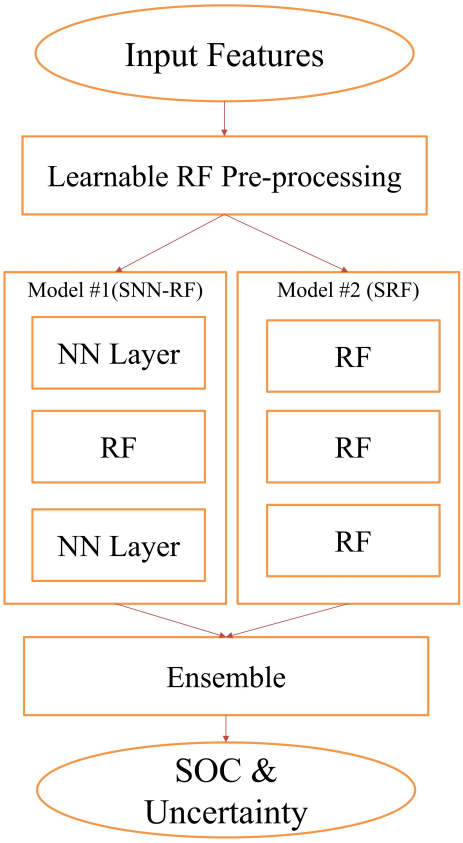

Figure 3. Schematic flowchart of ensemble framework.

Data Integration: Any missing values in the in situ SOC data were addressed using appropriate imputation techniques. Extensive ground sampling data, including SOC measurements, were utilized as ground truth for model training and validation, in addition to polarimetric SAR descriptors such as the $\mathbf{C}_2$ matrix, scattering entropy, and incidence angle. These descriptors provide valuable information about soil properties. SAR descriptors were normalized to ensure they were on a comparable scale. The in situ SOC data and SAR descriptors were integrated to form a comprehensive dataset. This integration ensured that the model was trained on a comprehensive dataset that captured both spatial and spectral variations.

Training: The dataset was split into 80% for training and 20% for independent validation. The training data were further stratified into 10 bins covering SOC values from 0.03% to 1.5% to ensure a diverse representation. Within this, an internal 80-20 split was used for model validation. The Random Forest model was configured with 200 trees, selected to balance computational efficiency and prediction stability with a size of 11,500 samples. The Gini impurity criterion was used for node splitting, and the maximum depth was set based

on cross-validation to prevent overfitting while capturing complex relationships in the data. For the SNN-RF model, the RF layer was trained on these stratified samples, learning non-linear relationships before passing outputs to the neural network for final prediction. In the Stacked RF model, multiple base RF models were trained independently, with their outputs fed into a meta-level RF model for final SOC estimation. This structured sampling approach ensured both models effectively captured SOC variability.

Validation: The model's performance was validated using the 20% independent validation dataset. Metrics such as R and RMSE, related to the sample size (N), were used to evaluate the accuracy and reliability of the predictions.

This methodology ensures a robust and accurate mapping of soil organic carbon percentage by leveraging the strengths of ensemble machine learning models and the rich information provided by ground and satellite data.

4. Results and Discussion

This section analyzes the predicted SOC values in the croplands of four districts. The spatial distribution of the SOC percentages within the croplands is illustrated in Figure 4. Although there is a trend in SOC percentage that appears to be consistent across all four districts, the limited in situ data distribution spatially does not allow for a statistically reliable comparison between the different agroclimatic zones.

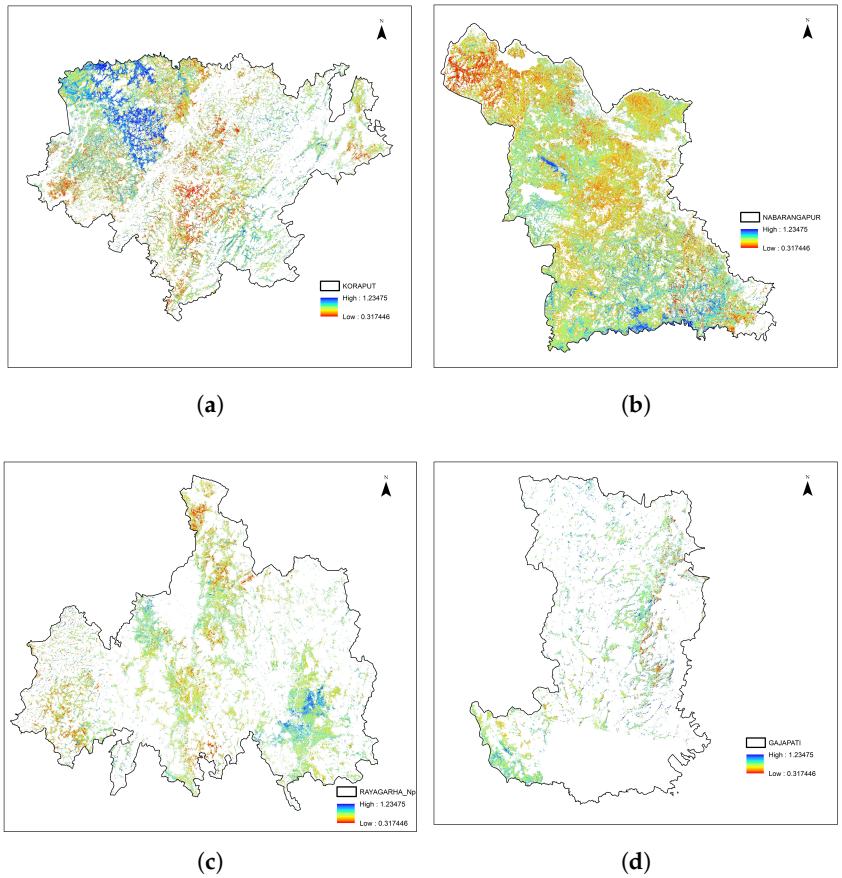

Figure 4. (**a**–**d**) Spatial distribution of SOC in each district with continuous range values.

We provide a detailed quantitative analysis of the four districts, examining the test sample size, correlation coefficient (r), RMSE, and standard error of the predicted values. The results are plotted in Figure 5. Koraput demonstrated the strongest agreement, with an r-value of 0.72 and an RMSE of 0.18, supported by a good sample size. This suggests a

reasonably strong predictive capability of the model in Koraput, although the scatter of points around the 1:1 line indicates some level of uncertainty in individual predictions, as reflected in the confidence intervals.

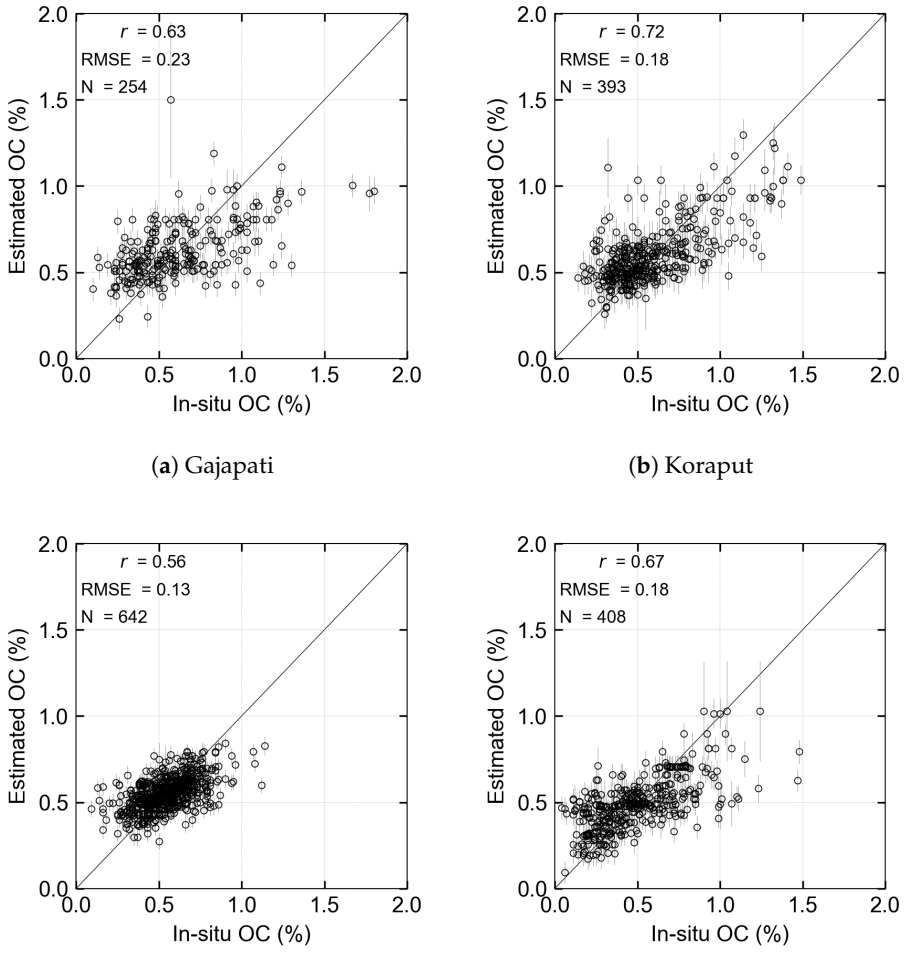

Figure 5. Correlation plots of measured SOC and estimated SOC with standard error for each district.

Crop types also play a crucial role in SOC distribution across the study area. In Koraput, which has the highest crop diversity, Niger (*Guizotia abyssinica*), maize (*Zea mays*), and finger millet (*Eleusine coracana*) are widely cultivated alongside rice (*Oryza sativa*). Nabarangpur features maize (*Zea mays*) and sugarcane (*Saccharum officinarum*) as additional major crops, while Rayagada and Gajapati districts are dominated by cotton (*Gossypium* spp.) and maize (*Zea mays*), with cotton being grown on newly converted agricultural lands. These variations in cropping patterns contribute to differences in SOC levels and may influence model performance across regions.

Rayagada exhibited a moderate positive correlation (r = 0.67), slightly lower than Koraput but still indicating a reasonable agreement between measured and estimated values. Gajapati had a lower correlation (r = 0.63), which could be attributed to a smaller sample size (N = 254) and the hilly terrain, introducing additional spatial variability. The weakest correlation was observed in Nabarangpur (r = 0.56). Despite its large sample size, the district's high cropping intensity and complex variations in soil management practices likely introduced challenges in capturing SOC variations effectively, leading to reduced model performance.

The overall accuracy of the four districts shows that averaging the large sample sizes of Nawrangpur and Rayagargh districts, compared to the slightly smaller Koraput and Gajapati districts, yields an accuracy r-value of 0.66 and an RMSE of 0.18%, as shown in Figure 6. The uncertainty analysis and accuracy assessment highlight the role of land use, cropping patterns, and agroclimatic conditions in SOC prediction accuracy. These findings underscore the need for further refinement of SOC estimation models to better account for localized agricultural and environmental factors.

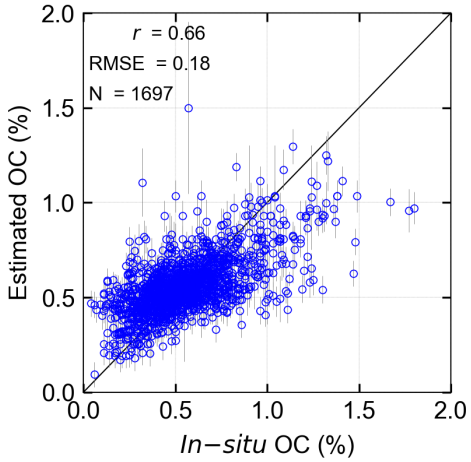

Figure 6. Correlation plot of measured SOC and estimated SOC for all districts.

The SOC distribution map generated from this study effectively captures the influence of key soil physical properties—drainage, topography, texture, and pH—on SOC variability in the Eastern Ghats of Odisha. The spatial patterns observed in our SOC estimations are consistent with established research findings, reinforcing the reliability of the produced map.

Well-drained soils exhibit higher SOC retention, whereas poorly drained or waterlogged areas hinder microbial activity and root interactions, leading to unstable SOC levels [29]. Conversely, excessive drainage accelerates organic matter decomposition, reducing long-term carbon storage [30]. This pattern is evident in the rice-dominated districts of Koraput and Nabarangpur, where intensive double-cropping practices coincide with lower SOC levels.

Soil texture further influences SOC distribution, with fine–loamy and fine–textured soils demonstrating higher carbon retention compared to coarser, loamy–skeletal soils. The higher silt and clay content in finer soils enhances SOC stabilization by promoting aggregate formation and reducing organic matter decomposition [31]. Similarly, neutral pH soils (~7) show greater SOC stability, while acidic soils—prevalent across approximately 70% of Odisha's cultivated land—exhibit lower SOC levels due to suppressed microbial activity and decomposition rates [32]. Topographic factors, particularly slope gradients, also impact SOC retention. Steeper slopes are prone to erosion, leading to the loss of organic matter, whereas gentle and very gentle slopes retain more SOC.

The generated SOC map has direct applications in sustainable land management and climate resilience strategies. By identifying areas with lower SOC levels, policymakers and farmers can prioritize soil conservation techniques to enhance carbon sequestration. Furthermore, integrating multi-frequency SAR data and analyzing the effects of different polarizations in SOC estimation could improve prediction accuracy, further aiding in carbon stock assessment.

5. Conclusions

This study integrates multi-frequency SAR data from Sentinel-1 (C-band) and ALOS-2/PALSAR-2 (L-band) with a large SOC sample size to assess spatial heterogeneity in the Eastern Ghats. The hybrid model demonstrated reliable performance (r = 0.64, RMSE = 0.18), highlighting the potential of machine learning and SAR data for the estimation of SOC. The unique agroclimatic conditions of the region, characterized by high rainfall and temperature, influence SOC distribution through vegetation cover, microbial activity, and erosion patterns.

Despite promising results, this study has limitations. The narrow dynamic range of in situ SOC data may affect model generalization, requiring smart sampling based on the initially produced SOC map. The resulting SOC map highlighted the influence of soil drainage, surface texture, and terrain on SOC distribution. The influence of crop types on SOC also needs deeper investigation with more balanced sampling.

In future research, we plan to explore advanced SAR missions like NISAR, incorporating multi-frequency and multi-polarization data for improved accuracy. Another avenue to refine SOC predictions is integrating additional spatial datasets (DEM, LULC, soil maps) and geostatistical modeling. Long-term monitoring of SOC changes under varying land-use and climatic conditions is crucial for developing sustainable soil management strategies.

Author Contributions: Conceptualization, P.K.B. and M.K.G.; methodology, P.K.B., M.K.G. and V.R.K.; software, P.K.B. and N.B.; validation, P.K.B. and N.B.; writing—original draft preparation, P.K.B. and N.B.; writing—review and editing, M.K.G. and V.R.K.; supervision, M.K.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors are grateful to CRAL, ICRISAT Patanchervu, for providing soil sample data. We would like to thank NITW, ICRISAT, for providing research facilities.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Adhikari, K.; Hartemink, A.E.; Minasny, B.; Bou Kheir, R.; Greve, M.B.; Greve, M.H. Digital mapping of soil organic carbon contents and stocks in Denmark. *PloS ONE* **2014**, *9*, e105519. [CrossRef] [PubMed]
2. Gerke, J. The central role of soil organic matter in soil fertility and carbon storage. *Soil Syst.* **2022**, *6*, 33. [CrossRef]
3. Lorenz, K.; Lal, R.; Ehlers, K. Soil organic carbon stock as an indicator for monitoring land and soil degradation in relation to United Nations' Sustainable Development Goals. *Land Degrad. Dev.* **2019**, *30*, 824–838. [CrossRef]
4. Batjes, N.H. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **2016**, *269*, 61–68. [CrossRef]
5. Scharlemann, J.P.; Tanner, E.V.; Hiederer, R.; Kapos, V. Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Manag.* **2014**, *5*, 81–91. [CrossRef]
6. Mulat, Y.; Kibret, K.; Bedadi, B.; Mohammed, M. Soil organic carbon stock under different land use types in Kersa Sub Watershed, Eastern Ethiopia. *Afr. J. Agric. Res.* **2018**, *13*, 1248–1256.
7. Zomer, R.J.; Bossio, D.A.; Sommer, R.; Verchot, L.V. Global sequestration potential of increased organic carbon in cropland soils. *Sci. Rep.* **2017**, *7*, 15554. [CrossRef] [PubMed]
8. Mandal, A.; Majumder, A.; Dhaliwal, S.; Toor, A.; Mani, P.K.; Naresh, R.; Gupta, R.K.; Mitran, T. Impact of agricultural management practices on soil carbon sequestration and its monitoring through simulation models and remote sensing techniques: A review. *Crit. Rev. Environ. Sci. Technol.* **2022**, *52*, 1–49. [CrossRef]
9. Doetterl, S.; Kearsley, E.; Bauters, M.; Hufkens, K.; Lisingo, J.; Baert, G.; Verbeeck, H.; Boeckx, P. Aboveground vs. belowground carbon stocks in African tropical lowland rainforest: Drivers and implications. *PloS ONE* **2015**, *10*, e0143209. [CrossRef]
10. Jian, J.; Du, X.; Reiter, M.S.; Stewart, R.D. A meta-analysis of global cropland soil carbon changes due to cover cropping. *Soil Biol. Biochem.* **2020**, *143*, 107735. [CrossRef]

11. Yu, Z.; Lu, C.; Cao, P.; Tian, H. Long-term terrestrial carbon dynamics in the Midwestern United States during 1850–2015: Roles of land use and cover change and agricultural management. *Glob. Change Biol.* **2018**, *24*, 2673–2690. [CrossRef]
12. Zeraatpisheh, M.; Garosi, Y.; Owliaie, H.R.; Ayoubi, S.; Taghizadeh-Mehrjardi, R.; Scholten, T.; Xu, M. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *Catena* **2022**, *208*, 105723. [CrossRef]
13. Rodionov, A.; Pätzold, S.; Welp, G.; Pude, R.; Amelung, W. Proximal field Vis-NIR spectroscopy of soil organic carbon: A solution to clear obstacles related to vegetation and straw cover. *Soil Tillage Res.* **2016**, *163*, 89–98. [CrossRef]
14. Chabrillat, S.; Ben-Dor, E.; Cierniewski, J.; Gomez, C.; Schmid, T.; van Wesemael, B. Imaging spectroscopy for soil mapping and monitoring. *Surv. Geophys.* **2019**, *40*, 361–399. [CrossRef]
15. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
16. Angelopoulou, T.; Tziolas, N.; Balafoutis, A.; Zalidis, G.; Bochtis, D. Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sens.* **2019**, *11*, 676. [CrossRef]
17. Petropoulos, G.P.; Ireland, G.; Barrett, B. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Phys. Chem. Earth Parts A/B/C* **2015**, *83*, 36–56.
18. Ottinger, M.; Kuenzer, C. Spaceborne L-band synthetic aperture radar data for geoscientific analyses in coastal land applications: A review. *Remote Sens.* **2020**, *12*, 2228. [CrossRef]
19. Camargo, F.F.; Sano, E.E.; Almeida, C.M.; Mura, J.C.; Almeida, T. A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sens.* **2019**, *11*, 1600. [CrossRef]
20. Odebiri, O.; Mutanga, O.; Odindi, J. Deep learning-based national scale soil organic carbon mapping with Sentinel-3 data. *Geoderma* **2022**, *411*, 115695. [CrossRef]
21. Santos, E.P.d.; Moreira, M.C.; Fernandes-Filho, E.I.; Demattê, J.A.M.; Dionizio, E.A.; Silva, D.D.d.; Cruz, R.R.P.; Moura-Bueno, J.M.; Santos, U.J.d.; Costa, M.H. Sentinel-1 imagery used for estimation of soil organic carbon by dual-polarization SAR vegetation indices. *Remote Sens.* **2023**, *15*, 5464. [CrossRef]
22. Zhou, Y.; Zhao, X.; Guo, X.; Li, Y. Mapping of soil organic carbon using machine learning models: Combination of optical and radar remote sensing data. *Soil Sci. Soc. Am. J.* **2022**, *86*, 293–310. [CrossRef]
23. Padarian, J.; Minasny, B.; McBratney, A.B. Machine learning and soil sciences: A review aided by machine learning tools. *SOIL* **2020**, *6*, 35–52. [CrossRef]
24. Suleymanov, A.; Tuktarova, I.; Belan, L.; Suleymanov, R.; Gabbasova, I.; Araslanova, L. Spatial prediction of soil properties using random forest, k-nearest neighbors and cubist approaches in the foothills of the Ural Mountains, Russia. *Model. Earth Syst. Environ.* **2023**, *9*, 3461–3471. [CrossRef]
25. Gumma, M.K.; Thenkabail, P.S.; Teluguntla, P.G.; Oliphant, A.; Xiong, J.; Giri, C.; Pyla, V.; Dixit, S.; Whitbread, A.M. Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google Earth Engine cloud. *GIScience Remote Sens.* **2020**, *57*, 302–322. [CrossRef]
26. Barakat, R. Degree of polarization and the principal idempotents of the coherency matrix. *Opt. Commun.* **1977**, *23*, 147–150. [CrossRef]
27. Dey, S.; Bhogapurapu, N.; Homayouni, S.; Bhattacharya, A.; McNairn, H. Unsupervised classification of crop growth stages with scattering parameters from dual-pol Sentinel-1 SAR data. *Remote Sens.* **2021**, *13*, 4412. [CrossRef]
28. Wiesmeier, M.; Barthold, F.; Blank, B.; Kögel-Knabner, I. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil* **2011**, *340*, 7–24. [CrossRef]
29. Kilpeläinen, J.; Peltoniemi, K.; Ojanen, P.; Mäkiranta, P.; Adamczyk, S.; Domisch, T.; Laiho, R.; Adamczyk, B. Waterlogging may reduce chemical soil C stabilization in forested peatlands. *Soil Biol. Biochem.* **2023**, *187*, 109229. [CrossRef]
30. Nachimuthu, G.; Hulugalle, N. On-farm gains and losses of soil organic carbon in terrestrial hydrological pathways: A review of empirical research. *Int. Soil Water Conserv. Res.* **2016**, *4*, 245–259. [CrossRef]
31. Cai, A.; Feng, W.; Zhang, W.; Xu, M. Climate, soil texture, and soil types affect the contributions of fine-fraction-stabilized carbon to total soil organic carbon in different land uses across China. *J. Environ. Manag.* **2016**, *172*, 2–9. [CrossRef] [PubMed]
32. Dash, P.K. Characterization, Taxonomic Classification and Fertility Status of Soils of Keonjhar District Under North Central Plateau Agro-Climatic Zone of Odisha. Ph.D. Thesis, Department of Soil Science and Agricultural Chemistry, OUAT, Bhubaneswar, India, 2022.