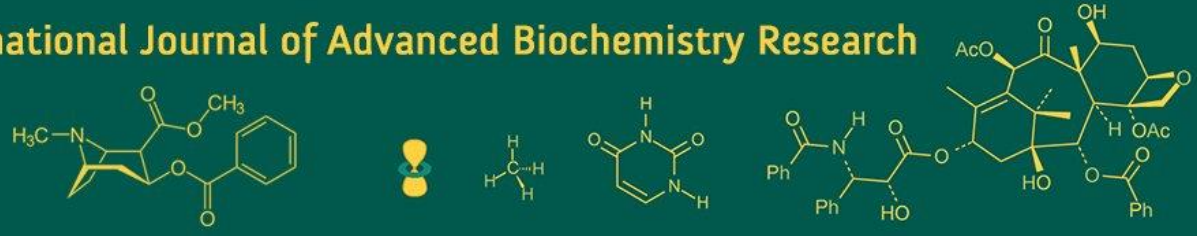
International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2025): 5.29
IJABR 2025; SP-9(10): 35-42
www.biochemjournal.com
Received: 08-08-2025
Accepted: 11-09-2025

G Tiwari
ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra, India

Dr. VN Mishra
Soil Science and Agricultural Chemistry Div., Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Dr. RP Sharma
Senior Scientist, ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Udaipur, Rajasthan, India

Dr. S Chattaraj
Scientist, ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Kolkata, Maharashtra, India

Dr. B Dash
Scientist, ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra, India

A Jangir
Scientist, ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Udaipur, Rajasthan, India

Dr. LC Malav
Scientist, ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Udaipur, Rajasthan, India

Corresponding Author:
G Tiwari
ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, Maharashtra, India

Soil organic carbon mapping in India's black soil region through machine learning

G Tiwari, VN Mishra, RP Sharma, S Chattaraj, B Dash, A Jangir and LC Malav

DOI: https://www.doi.org/10.33545/26174693.2025.v9.i10Sa.5832

Abstract
Soil organic carbon (SOC) is a critical component of soil health, influencing productivity, ecosystem services, and carbon sequestration. This study applied Quantile Random Forest (QRF) modeling to predict SOC distribution across depth intervals (0-200 cm) using 88 soil profiles from India's Black Soil Region (BSR). Observed, splined, and predicted SOC values were compared to evaluate prediction accuracy and spatial variability. Results indicate low SOC levels (mean = 0.5%) consistent with semi-arid agroecosystems, with highest concentrations in surface soils (0-15 cm) and a decline with depth. Prediction performance improved with depth (R^2 = 0.48 at 0-5 cm; 0.62 at 100-200 cm), reflecting greater stability of SOC in subsoils. Vegetation, slope, rainfall, and temperature emerged as key predictors of SOC distribution. Spatial mapping highlighted higher SOC in forested and hilly zones, and lower SOC in cultivated and degraded lands. Uncertainty analysis revealed greater variability in surface layers than subsoils. These findings provide insights into SOC dynamics in black soils and demonstrate the utility of machine learning for depth-resolved SOC prediction, supporting sustainable land management and carbon conservation.

Keywords: Digital soil mapping, prediction intervals, machine learning, Scorpan, Deccan trap, spatial variability, land use planning

Introduction
Soil organic carbon (SOC) is a fundamental attribute of soils, regulating fertility, productivity, and ecosystem services. It enhances soil structure, water-holding capacity, nutrient cycling, and cation exchange, thereby sustaining agricultural production (Lal, 2004) [11]. Beyond agronomic benefits, SOC is the largest terrestrial carbon pool, storing more carbon than the atmosphere and vegetation combined, and thus plays a central role in climate-change mitigation through carbon sequestration (Batjes, 1996) [1]. Consequently, understanding and accurately predicting SOC stocks and distributions is critical for food security, soil health, and climate resilience.

Globally, SOC distribution is highly uneven both horizontally and vertically. Surface horizons contain the majority of labile SOC due to organic matter inputs from roots and crop residues, while subsoils store more stable yet often smaller fractions of SOC (Jobbágy & Jackson, 2000; Rumpel & Kögel-Knabner, 2011) [8, 23]. Climate, vegetation, parent material, and land use interact to regulate SOC accumulation and decomposition, leading to pronounced variability across landscapes (Wiesmeier *et al*., 2019) [30]. In India, SOC stocks are generally low compared to global averages, largely due to semi-arid climates, intensive cultivation, residue removal, and limited organic matter inputs (Bhattacharyya *et al*., 2000) [24]. These conditions make regional SOC prediction especially challenging.

Digital soil mapping (DSM) has emerged as a promising framework for SOC prediction by integrating soil profile data with environmental covariates derived from terrain analysis, climate models, and remote sensing. Machine learning (ML) algorithms such as Random Forest (RF), Cubist, and Support Vector Machines (SVM) have been widely adopted in DSM due to their ability to capture complex non-linear interactions between SOC and environmental drivers (Hengl *et al*., 2017) [7]. However, a key limitation of many ML models is their inability to adequately quantify uncertainty, which is critical for risk-aware decision-making in soil management (Ma *et al*., 2014) [16].

This challenge is particularly pronounced in India's Black Soil Region (BSR), dominated by Vertisols and associated soils. The BSR is a major agricultural zone that supports staple crops such as cotton, soybean, and sorghum, yet it is characterized by semi-arid climates, intensive land use, and inherently low SOC stocks (Bhattacharyya *et al*., 2007) [2]. High spatial heterogeneity in land use, soil depth, and management further complicates SOC prediction. Moreover, field-based SOC datasets are sparse and unevenly distributed, limiting the applicability of conventional statistical models and highlighting the need for robust ML methods that can handle data scarcity while quantifying prediction uncertainty.

To address this, Quantile Regression Forest (QRF) offers a suitable approach. QRF is an extension of RF that predicts conditional quantiles, providing not only central predictions but also prediction intervals without distributional assumptions (Meinshausen, 2006) [18]. By retaining the full distribution of response values within decision trees, QRF captures both mean trends and variability, thereby enabling uncertainty quantification. This makes it especially valuable for SOC prediction in regions like the BSR, where sample density is low and environmental heterogeneity is high. Therefore, the objectives of this study are to: (i) Harmonize profile SOC data across standard depth intervals (0-200 cm) using mass-preserving splines; (ii) Develop and validate QRF models for predicting SOC distributions using environmental covariates; (iii) Compare observed, splined, and predicted SOC values to evaluate model accuracy; and (iv) Quantify prediction uncertainty and generate depth-resolved SOC maps to support sustainable land management and carbon accounting in India's Black Soil Region

2. Materials and Methods

2.1. Study area

The study was conducted in the BSR of Amravati district of Maharashtra, India, covering approximately 59,758 ha (Fig. 1) and lies between 20○24' to 21○33'N and 77○06' to 78○18'E. The region is part of the Deccan Plateau, characterized by flat-topped hills (plateaus) and intervening valleys. The climate is semi-arid tropical with a mean annual rainfall of 975 mm and a mean annual temperature of 28 °C. The geology is predominantly Deccan Trap basalt, with alluvial deposits in the valley of the Purna River. Rainfed agriculture is the dominant land use. The soils have an ustic moisture regime and an isohyperthermic temperature regime.

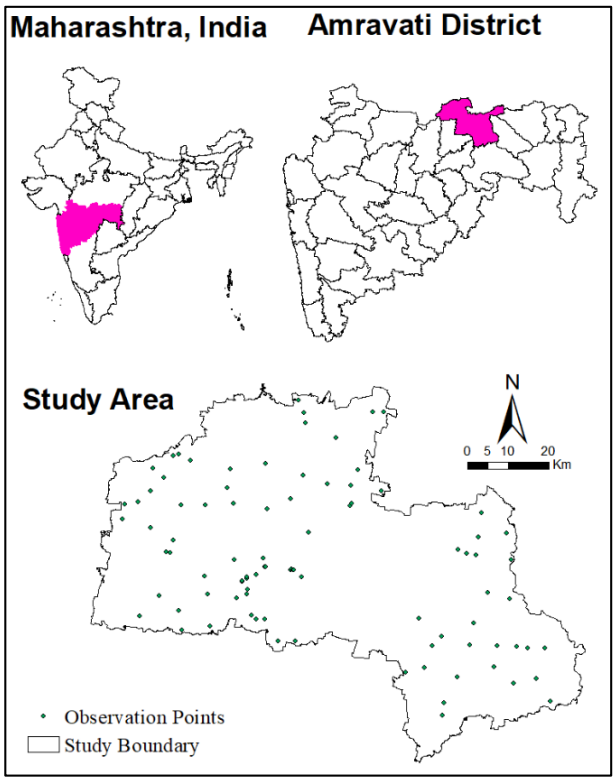


Fig 1: Location of study area

2.2 Soil data

A total of 88 geo-referenced soil profiles, with depths ranging from 14 to 155 cm, were used in this study. Profiles were sampled following standard protocols, and laboratory analyses were performed to determine SOC concentration using wet oxidation (rapid titration) method. Ground soil sample passed through a 0.5 mm sieve were used for estimating organic carbon. Soil samples were oxidised by potassium dichromate (1 N) and the conc. H_2SO_4 was used to generate the heat of dilution. The amount of dichromate unutilized was determined by back titration with standard ferrous ammonium sulphate solution (0.5 N). To ensure comparability across profiles with unequal depth intervals, SOC data were harmonized to standardized depth intervals (0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm) using the mass-preserving spline approach. This procedure allowed the integration of profile data into a continuous depth function while conserving SOC mass.

2.3 Environmental covariates

Based on the SCORPAN framework, 47 covariates representing Soil, Climate, Organisms, Relief, Parent material, Age, and Space were compiled (Table 1). Terrain attributes (30 m resolution) were derived from the SRTM

DEM using SAGA GIS. Climate variables (Mean Annual Precipitation and Temperature) were sourced from *WorldClim* (1km resolution). Time-series Landsat 5 TM imagery was used to compute spectral indices (NDVI, EVI, SAVI, FV) for three seasons (*kharif*, *rabi*, *zaid*). Annual average Land Surface Temperature (LST) was derived from MODIS data (1km). All covariates were resampled to a 30m grid using bilinear interpolation.

Table 1: Environmental covariates used for digital soil mapping of SOC

S. N	Group	Covariate	Abbr.	Res.
1	Climate	Mean annual precipitation (mm)	MAP	1 km
2	Terrain	Elevation (m)	Elv	30 m
		Slope (%)	Slope	30 m
		Relative Slope Position	RSP	30 m
		Channel Network Base Level	CNBL	30 m
		Channel Network Distance	CND	30 m
		Multi-Resolution Ridge Top Flatness Index	MRRTF	30 m
		Multi-Resolution Valley Bottom Flatness Index	MRVBF	30 m
		Valley Depth	VD	30 m
		Topographic Wetness Index	TWI	30 m
		LS-Factor	LSf	30 m
3	Vegetation (*Kharif* (k), *Rabi* (r), *Zaid* (z))	Land surface thermal conditions	LST	30 m
		Normalized Difference Vegetation Index.	NDVI	10 m
		Near infrared	NIR	10 m
		Enhanced Vegetation Index	EVI	10 m
		Fractional vegetation	FV	10 m

2.4 Modelling framework

To avoid overfitting and reduce multicollinearity, recursive feature elimination (RFE) was performed using the RFE function in the caret R package. Variables were ranked by their importance (%IncMSE) from a preliminary RFE model, and the optimal subset that maximized model performance was selected.

The QRF model was implemented using the ranger package in R. The model was tuned via out-of-bag (OOB) error estimation; the optimal parameters were *mtry* = 5, *num.trees* = 1000, and *min.node.size* = 5. Unlike standard RF, which estimates the conditional mean, QRF retains the entire distribution of values in the leaf nodes of each tree, allowing for the computation of any quantile of the conditional distribution. This study generated the 0.05, 0.5 (median), and 0.95 quantiles to represent the lower bound, median prediction, and upper bound of the 90% prediction interval, respectively.

2.5 Model validation

Model performance was evaluated using a repeated (20 times) 10-fold cross-validation. Performance metrics included coefficient of determination (R^2), root mean square error (RMSE), mean error (ME), and Lin's concordance correlation coefficient (CCC). Good models have a root mean square error that is close to 0, R^2 and CCC that is equal to or close to 1.

$$\text{Coefficeint of determination } (R^2) = 1-\frac{\sum_{i=1}^{n}(\text{pi}-\text{oi})^2}{\sum_{i=1}^{n}(\bar{\text{pi}}-\bar{\text{oi}})^2} \quad \text{(i)}$$

$$\text{Mean error (ME)} = \frac{1}{n}\sum_{i=1}^{n}(\text{pi}-\text{oi}) \quad \text{(ii)}$$

$$\text{Root mean squared error (RMSE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{pi}-\text{oi})^2} \quad \text{(iii)}$$

where, pi and oi are predicted and observed values, $\bar{\text{pi}}$ and $\bar{\text{oi}}$ are means of these values.

$$\text{Lin's concordance correlation coefficient (CCC)} = \frac{2\rho\sigma_0\sigma_p}{{\sigma_0}^2+{\sigma_p}^2+(\mu_0-\mu_p)^2} \quad \text{(iv)}$$

In this formula, ρ is the Pearson correlation coefficient between the observed and predicted values, μ_0 and μ_p are the means of the observed and predicted values, and ${\sigma_0}^2$ and ${\sigma_p}^2$ are the corresponding variances.

2.6 Uncertainty quantification

A key advantage of QRF is its capacity to estimate prediction intervals. For each grid cell, the 5th and 95th quantiles of the SOC distribution were extracted, and the prediction interval width (PIW) was computed as the difference between them. PIW maps were generated to spatially represent prediction uncertainty across the study area. Narrow PIWs indicate greater model confidence, whereas wider PIWs highlight regions where SOC predictions are less reliable and where additional sampling may be required.

3. Results and Discussions

3.1 Observed SOC

Observed SOC values (14-155 cm) indicate a wide range (0.1-2.2%), with a mean of 0.5% and a SD of 0.3 (Table 2). The CV = 0.6 suggests moderate variability in SOC content within the observed data. The sample size (N) for this analysis is 88, to train the model for soil SOC and which provides a reasonable basis for statistical interpretation. The observed mean SOC value of 0.5% in this study is relatively low compared to global averages, which often range between 1.5% and 3.5% in temperate and tropical regions (Lal, 2004) [11]. However, the low SOC content observed here is consistent with findings in arid and semi-arid regions, where limited vegetation covers and high temperatures accelerate organic matter decomposition (Batjes, 1996) [1]. The studies in similar climatic zones, such as those in sub-Saharan Africa and parts of Australia, have reported SOC values ranging from 0.2% to 1.5%.

In India, SOC levels vary significantly across regions due to differences in climate, land use, and soil management

practices. The observed SOC range (0.1% to 2.2%) in this study aligns with findings from other parts of India, particularly in regions with similar agro-climatic conditions. The studies in the Deccan Plateau and semi-arid regions of Maharashtra have reported SOC values ranging from 0.3% to 2.5% (Bhattacharyya *et al*., 2007) [2]. The low mean SOC value (0.5%) in this study may reflect the impact of intensive agriculture, limited organic matter inputs, and high rates of soil erosion, which are common challenges in Indian soils. Studies in Maharashtra have reported SOC values ranging from 0.4% to 2.8% in the topsoil, with lower values in rainfed and degraded lands.

3.2 Soil SOC after splining

The depth-wise distribution of SOC, harmonized using mass-preserving splines, revealed a clear decline with increasing soil depth. The surface layer (0-15 cm) recorded the highest mean SOC concentration of 0.7%, while values steadily decreased with depth, reaching 0.4% in the 100-200 cm interval. This vertical gradient is consistent with the widely documented global pattern, where SOC stocks are concentrated in surface horizons due to greater organic matter inputs from litter deposition and root turnover (Jobbágy & Jackson, 2000; Wiesmeier *et al*., 2019) [8, 30]. Although SOC content diminishes with depth, the subsoil still represents an important reservoir for long-term carbon storage because of its slower turnover rates and relative protection from microbial decomposition (Rumpel & Kögel-Knabner, 2011; Luo *et al*., 2019) [23, 15]. These results emphasize that while management interventions often focus on topsoil carbon, deeper soil layers must also be considered in carbon accounting and climate mitigation strategies.

Table 2: Descriptive statistics of observed, splined and predicted soil SOC

Depth (cm)	Min	Max	Mean	SD	CV%	N
Observed SOC						
14-155	0.1	2.2	0.5	0.3	0.6	88
SOC after splining						
0-5	0.17	1.9	0.7	0.3	0.4	84
5-15	0.16	1.8	0.7	0.3	0.4	84
15-30	0.13	1.5	0.6	0.2	0.3	79
30-60	0.01	1.5	0.5	0.2	0.4	70
60-100	0.03	1.8	0.5	0.3	0.6	66
100-200	0.03	1.7	0.4	0.3	0.8	57
Predicted SOC						
0-5	0.17	1.91	0.71	0.29	0.4	84
5-15	0.48	0.87	0.68	0.09	0.1	83
15-30	0.45	0.83	0.66	0.09	0.1	83
30-60	0.39	0.95	0.59	0.11	0.2	80
60-100	0.27	0.63	0.49	0.10	0.2	70
100-200	0.14	0.59	0.37	0.12	0.3	57

3.3 Variables of importance

The Quantile Regression Forest (QRF) model identified distinct sets of controlling factors for SOC distribution across soil depths. In the surface layers, vegetation density (VD), slope, and remote sensing-derived indices emerged as the most influential predictors, reflecting the strong influence of biomass inputs, land cover, and topographic control on organic matter accumulation. At subsurface depths, climatic variables such as mean annual precipitation (MAP) and land surface temperature (LST) became more critical, highlighting the role of soil-climate interactions in regulating carbon stabilization below the plough layer. In the deeper horizons (100-200 cm), terrain attributes, particularly multi-resolution valley bottom flatness (MRVBF) and multi-resolution ridge top flatness (MRRTF), gained importance, indicating the influence of long-term geomorphic processes on SOC distribution. These findings are consistent with digital soil mapping (DSM) studies in India and globally, which emphasize the depth-dependent influence of biotic, climatic, and terrain factors on SOC variability (Hengl *et al*., 2017; Wadoux *et al*., 2020) [7, 29].

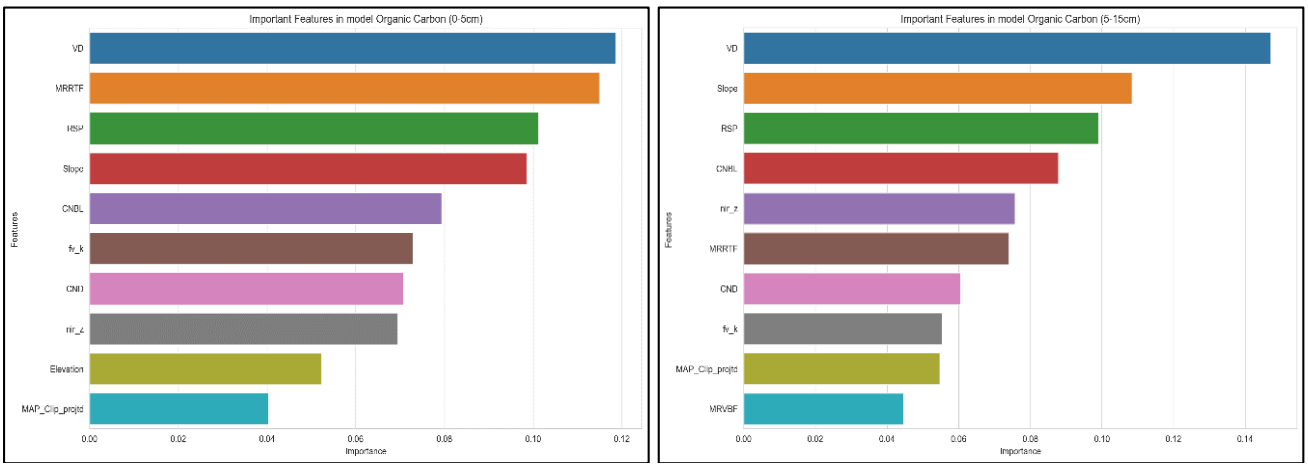


Fig 2: Important variable for predicting soil SOC

3.4 Predicted SOC and spatial distribution

The QRF-based predictions of SOC ranged from 0.04 to 1.2% across the study area (Fig. 3). Similar to the observed and splined data, a distinct depth-dependent decline was evident, with the highest concentrations in the surface horizons and progressively lower values with depth. However, the predicted SOC variability was generally lower than the observed values, particularly in the topsoil, reflecting the model's tendency to smooth local heterogeneity. Spatial maps revealed clear geographic patterns, with higher SOC levels concentrated in forested and hilly areas, where organic inputs are greater and erosion is less severe, while lower values were associated with intensively cultivated or degraded zones. These spatial patterns are consistent with regional studies in India (Patil *et al*., 2020) [21] and global assessments of SOC distribution (Lugato *et al*., 2014) [14], which similarly highlight the strong influence of land use, topography, and management intensity on SOC variability.

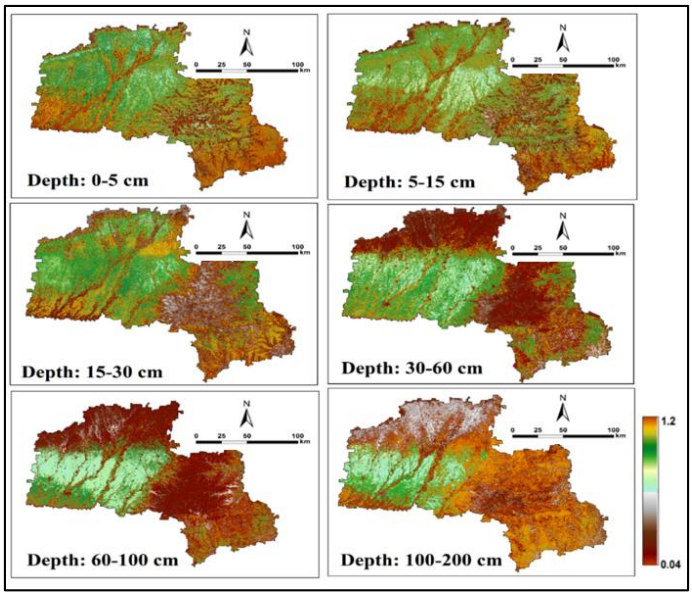


Fig 3: Depth-wise predicted map for the soil SOC

The spatial distribution of weighted mean of SOC in the study area is illustrated in Fig. 4, SOC levels range from 0.2% to 0.7%. The spatial variability suggests that higher SOC is found in hilly or forested regions, while lower SOC is prevalent in agricultural or degraded lands. SOC distribution varies significantly across different landscapes due to factors such as climate, vegetation, topography, and land use. High SOC areas, primarily in the western parts of the region, are associated with Purna valley deposition. These areas benefit from higher rainfall, minimal soil disturbance, and organic matter accumulation, similar to findings in the Western Ghats, Himalayan forests, and central India's dense forest regions. Moderate SOC areas are found in transitional zones between forests and agricultural lands, often characterized by agroforestry, mixed cropping, or semi-intensive cultivation that maintains SOC at intermediate levels.

This pattern is also observed in Madhya Pradesh, Chhattisgarh, and Odisha, where forest-agriculture boundaries significantly influence SOC levels. Low SOC areas, mostly in the southern and eastern parts of the region, are likely associated with intensively cultivated, semi-arid, or degraded lands. In these areas, high oxidation rates, deforestation, overgrazing, and soil erosion contribute to SOC depletion, a trend also reported in Maharashtra's Vidarbha and Marathwada regions, Rajasthan's drylands, and Gujarat's arid zones (Patil *et al*., 2020b) [21].

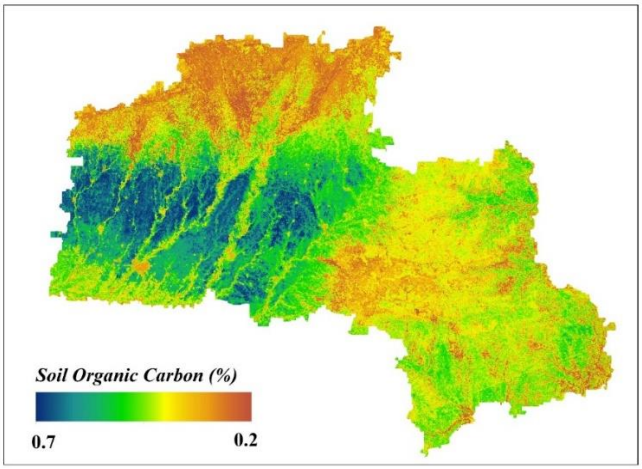


Fig 4: Predicted map for the weighted mean of SOC

3.5 Prediction Accuracy (QRF Model)

The predictive performance of the QRF model varied with soil depth, showing an overall improvement in accuracy for deeper layers (Table 3). In the shallow surface soils (0-5 cm), the model achieved an R^2 of 0.48, reflecting the challenges of capturing the high spatial variability of topsoil SOC, which is strongly influenced by land management, residue inputs, and localized erosion-deposition processes. In contrast, prediction accuracy improved progressively with depth, reaching an R^2 of 0.62 in the 100-200 cm interval. This enhanced performance at greater depths can be attributed to the relatively stable nature of subsoil carbon stocks, which are less affected by short-term anthropogenic disturbances and better explained by long-term soil-forming factors. Such depth-dependent trends in predictive performance are consistent with global DSM studies, which have also reported stronger model fits for subsoil SOC compared to surface layers (Zhao *et al*., 2021; Kang *et al*., 2022) [34, 9].

Table 3: Results of QRF model validation for SOC

Depth (cm)	RMSE	R^2	CCC
0-5	0.19	0.48	0.54
5-15	0.16	0.53	0.60
15-30	0.14	0.58	0.67
30-60	0.11	0.61	0.70
60-100	0.12	0.57	0.67
100-200	0.12	0.2	0.72

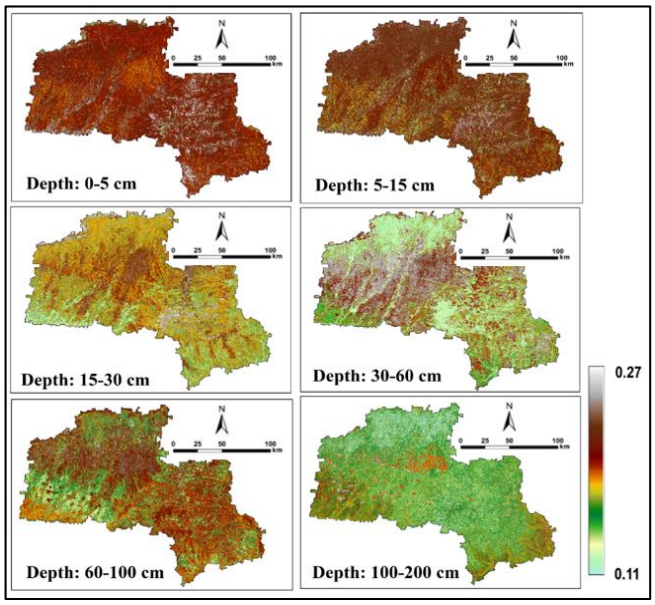


Fig 5: Depth-wise predictive uncertainty map for soil SOC

3.6 Uncertainty Analysis

The uncertainty analysis revealed a clear depth-dependent pattern, with higher uncertainty in the shallow layers (0-15 cm) and progressively lower uncertainty in the subsoil horizons (Fig 5). This trend reflects the inherently greater variability of surface SOC, which is strongly influenced by land management, residue turnover, and localized erosion-deposition processes, compared to the more stable carbon pools at depth. The QRF model effectively quantified prediction intervals, offering a robust representation of uncertainty without assuming a predefined error distribution. These findings are consistent with global SOC mapping initiatives such as SoilGrids (Hengl *et al*., 2017) [7], where surface predictions are typically associated with wider confidence intervals. In the Indian black soil regions, the challenge of capturing surface variability is particularly pronounced due to intensive cultivation and semi-arid climatic conditions, while deeper soils display more predictable carbon stocks. This highlights the importance of explicitly accounting for uncertainty in SOC predictions to support risk-aware land management and climate mitigation strategies.

4. Conclusion

This study provides a robust framework for predicting soil organic carbon (SOC) in India's Black Soil Region (BSR) using digital soil mapping and machine learning. The QRF model effectively captured depth-dependent SOC dynamics, delivering higher accuracy in subsoils and quantifiable uncertainty across the profile. Results confirm that vegetation and land use drive surface SOC variability, while climate and terrain govern deeper pools. The final SOC maps highlight the impact of cultivation and degradation on carbon depletion and identify forested and hilly zones as

regional carbon hotspots. These outputs serve as decision-support tools for sustainable soil management and climate change mitigation in semi-arid black soils. Future research should integrate denser sampling and proximal sensing data to refine topsoil predictions and extend the framework across India's broader black soil belt.

5. Acknowledgment
I am deeply grateful to the faculty of the Soil Science Division, IGKV, Raipur, for their invaluable academic guidance. My sincere thanks to the faculty of the SRS Division and Director of ICAR-NBSS & LUP, for their expert mentorship and for providing an exceptional research environment.

References

1. Batjes NH. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma. 2016;269:61-68.
2. Bhattacharyya T, Pal DK, Mandal C, Velayutham M, Chandran P. Organic carbon stock of Indian soils and their geographical distribution. Current Science. 2007;91(5):611-616.
3. Catani F, Segoni S, Falorni G. An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. Water Resour Res. 2010;46(5):W05514.
4. Cheng Q, Li H, Wang J, Zhu AX. Soil property mapping using machine learning methods: A case study of regional soil organic matter content in China. Geoderma. 2019;333:100-109.
5. Gu H, Zhu AX, Liu J, Qin CZ, Zhou C. Soil thickness mapping using environmental correlation and geostatistical techniques in hilly areas of China. Soil Sci Soc Am J. 2018;82(3):609-620.
6. Guo L, Gong P, Amundson R, Yu Q. Quantifying the impact of soil data quality on digital soil mapping. Geoderma. 2019;337:93-103.
7. Hengl T, de Jesus JM, Heuvelink GBM, Gonzalez MR, Kilibarda M, Blagotić A, *et al*. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One. 2017;12(2):e0169748.
8. Jobbágy EG, Jackson RB. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol Appl. 2000;10(2):423-436.
9. Kang J, Zhang G, Liu F, Yang R. Depth-dependent prediction of soil organic carbon using machine learning: Implications for soil carbon modeling. Geoderma. 2022;405:115424.
10. Lagacherie P, Arrouays D, Bourennane H, Gomez C, Boulonne L. Digital soil mapping: Towards worldwide soil information. In: Sparks DL, editor. Advances in Agronomy. Vol. 157. Cambridge (MA): Academic Press; 2019. p.1-52.
11. Lal R. Soil carbon sequestration impacts on global climate change and food security. Science. 2004;304(5677):1623-1627.
12. Liu F, Zhang GL, Song XD, Li DC, Zhao YG. Pedogenesis-related soil thickness mapping of hilly areas in subtropical China. Soil Tillage Res. 2013;133:40-47.
13. Liu F, Xu X, Li D, Pan X, Zhang G. Mapping soil thickness using environmental variables and random forest. Geoderma. 2019;354:113859.
14. Lugato E, Panagos P, Bampa F, Jones A, Montanarella L. A new baseline of organic carbon stock in European agricultural soils using a modelling approach. Glob Change Biol. 2014;20(1):313-326.
15. Luo Z, Wang E, Sun OJ. Soil carbon dynamics and climate change: Current understanding and future challenges. Earth Sci Rev. 2019;196:102873.
16. Ma Y, Minasny B, McBratney AB. Uncertainty analysis for soil property mapping using random forest with quantile regression. Geoderma. 2014;232-234:243-252.
17. Mandal C, Bhattacharyya T, Sarkar D. Soil organic carbon stock in India and its significance. In: Sarkar D, Minhas PS, Singh R, editors. Soil Organic Carbon: A Sustainability Indicator for Soil Productivity and Health. Singapore: Springer; 2020. p.251-276.
18. Meinshausen N. Quantile regression forests. J Mach Learn Res. 2006;7:983-999.
19. Meyer WB, Turner BL, Skole D. Human drivers of global land-use change: Potentials of integrated assessment. In: Proceedings of the International Geosphere-Biosphere Programme. 2007. p.395-411.
20. Mulder VL, Lacoste M, Richer-de-Forges AC, Arrouays D. GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meters depth. Sci Total Environ. 2016;573:1352-1369.
21. Patil S, Kumar S, Singh SK, Chandran P. Spatial distribution of soil organic carbon in relation to land use and management practices in Central India. Catena. 2020;189:104467.
22. Pelletier JD, Rasmussen C. Geomorphically based predictive mapping of soil thickness in upland watersheds. Water Resour Res. 2009;45(9):W09417.
23. Rumpel C, Kögel-Knabner I. Deep soil organic matter-A key but poorly understood component of terrestrial C cycle. Plant Soil. 2011;338(1-2):143-158.
24. Singh SK, Sarkar D, Bhattacharyya T. Digital soil mapping in India: Status and prospects. Current Science. 2020;118(12):1866-1876.
25. Sreenivas K, Das BS, Mohanty BP, Wani SP. Spatial variability of soil organic carbon stocks in a semi-arid tropical watershed. Land Degrad Dev. 2016;27(7):1718-1728.
26. Tesfa TK, Tarboton DG, Chandler DG, McNamara JP. Modeling soil depth from topographic and land cover attributes. Water Resour Res. 2009;45(10):W10438.
27. Vågen T-G, Lal R, Singh BR. Soil carbon sequestration in sub-Saharan Africa: A review. Land Degrad Dev. 2016;27(3):574-589.
28. Vanwalleghem T, Stockmann U, Minasny B, McBratney AB. A quantitative model for integrating landscape evolution and soil formation. J Geophys Res Earth Surf. 2010;115(F4):F04019.
29. Wadoux AMJC, Minasny B, McBratney AB, Malone BP. Machine learning for digital soil mapping: Applications, challenges, and future directions. Comput Electron Agric. 2020;173:105379.
30. Wiesmeier M, Urbanski L, Hobley E, Lang B, von Lützow M, Marin-Spiotta E, *et al*. Soil organic carbon storage as a key function of soils-A review of drivers and indicators at various scales. Geoderma. 2019;333:149-162.

31. Yan X, Cai C, Fang H. Spatial prediction of soil depth using ordinary kriging and auxiliary variables in subtropical China. Geoderma. 2021;383:114745.
32. Yang R, Zhang G, Liu F. Depth functions and pedogenetic modeling of soil thickness. Soil Sci Soc Am J. 2020;84(3):806-817.
33. Zhang G, Yang R, Liu F. Predictive mapping of soil depth using random forest and Cubist in complex terrain. Catena. 2021;200:105125.
34. Zhao Y, Zhu AX, Qin CZ, Li B. Prediction of soil organic carbon at multiple depths: Comparison of machine learning models and geostatistical approaches. Geoderma. 2021;385:114874.