



Modelling Soil Organic Carbon Across a Rugged Altitudinal Gradient in Drylands: A Machine Learning Approach Using Multi-Source Covariates

Tumuzghi Tesfay^{1,2}, Nazih Y. Rebouh^{1,3}, Dmitry E. Kucher¹,
Balwan Singh², Elsayed S. Mohamed^{1,4}, Igor Yu. Savin^{1,3}

10.18805/IJARE.AF-1030

ABSTRACT

Background: Soil organic carbon (SOC) is essential for soil health, food security and climate change mitigation. However, reliable data on SOC distribution and its environmental drivers remain limited in data-scarce dryland regions like Eritrea. This hinders effective soil management and restoration planning.

Methods: SOC modelling was conducted across an altitudinal gradient landscape using environmental covariates and machine learning. Three predictor sets (46, 28 and 11 variables) were used selected through three approaches: 1) no selection, 2) removal of highly collinear ($r \geq 0.90$) and non-significant variables and 3) boruta algorithm-based selection. The predictive performance of cubist, random forest (RF) and partial least squares (PLS) algorithms were evaluated.

Result: SOC levels across the study area were generally low (mean = 0.71%). Rainfed croplands and communal grazing areas showed particularly depleted SOC, attributed to unsustainable land management, while forest and irrigated systems retained significantly higher SOC, indicating greater carbon sequestration potential. The Cubist model with 46 predictors performed best ($R^2 = 0.7465$, RPD = 2.0895), whereas PLS with 11 variables had the lowest accuracy ($R^2 = 0.5930$, RPD = 1.6491). Temperature emerged as the strongest predictor, followed by land use, altitude, Soil Organic Carbon Index, Landsat 8 band B10 and rainfall. The dominance of temperature for SOC prediction was supported by the strong negative correlation of SOC with temperature ($r = -0.582$) and positive with altitude ($r = 0.580$). These underscore the role of climate on the spatial-temporal dynamics of SOC and highlight for climate-smart strategies. Thus, we conclude that cost-effective assessments and monitoring of SOC that support evidence-based strategies for enhancing soil health, land restoration and climate resilience are possible through the developed Cubist and RF models from the Eritrean Central Highlands to the Western Midlands and similar environments.

Key words: Climate-smart agriculture, Land use, Machine learning, Temperature, Topography.

INTRODUCTION

Soil organic carbon (SOC) is a central driver of soil health and a key component of the global carbon cycle (FAO, 2017). How soils are managed determines whether they sequester or emit carbon. When managed well, SOC enhances nutrient cycling, soil structure and microbial diversity, while supporting ecosystem restoration and climate change mitigation (Page *et al.*, 2020). Conversely, poor management leads to carbon loss through emissions and nutrient mining. Thus, assessment, improvement and monitoring of SOC are essential for the well being of soils, ecosystems, societies and the planet.

Eritrea is found in the drought prone Sub-saharan Sahel region and more than 72% of its territory experiences arid to semi desert climatic conditions (Tesfay *et al.*, 2025a). Reports indicate widespread land degradation, desertification, recurring droughts, rising temperatures and declining rainfall over the past decades (MoA, 2018; Ghebregabher *et al.*, 2019; Tesfay *et al.*, 2024). Over 75% of the population relies on rainfed cropping and livestock rearing (FAO, 2021; Tesfay *et al.*, 2024) though soil fertility remains poor and crop yields are critically low; frequently below 0.7 t ha⁻¹ (Tesfay *et al.*, 2018).

¹Department of Environmental Management, Institute of Environmental Engineering, RUDN University, 6 Miklukho-Maklaya St., 117198 Moscow, Russia.

²National Higher Education and Research Institute, Hamelmalo Agricultural College, Keren, Eritrea.

³V.V. Dokuchaev Soil Science Institute, Pyzhevsky per. 7, Building 2, 119017 Moscow, Russia.

⁴National Authority for Remote Sensing and Space Sciences, Cairo 1564, Egypt.

Corresponding Author: Tumuzghi Tesfay, Department of Environmental Management, Institute of Environmental Engineering, RUDN University, 6 Miklukho-Maklaya St., 117198 Moscow, Russia. Email: tumuzghitesfay7@mail.com

ORCID: 0000-0002-0771-5522, 0000-0002-8621-6595, 0000-0001-5703-4621, 0000-0002-8739-5441

How to cite this article: Tesfay, T., Rebouh, N.Y., Kucher, D.E., Singh, B., Mohamed, E.S. and Savin, I.Y. (2026). Modelling Soil Organic Carbon Across a Rugged Altitudinal Gradient in Drylands: A Machine Learning Approach Using Multi-Source Covariates. *Indian Journal of Agricultural Research*. **60(6)**: 926-935. doi: 10.18805/IJARE.AF-1030.

Submitted: 17-02-2026 **Accepted:** 14-04-2026 **Online:** 29-04-2026

The absence of reliant soil information limits the capacity to plan and monitor improvements in soil fertility, evaluate the effectiveness of land rehabilitation efforts and support evidence based land management decisions. Soil studies in the country are scarce and predominantly rely on conventional coring and laboratory analysis, which are labour intensive, costly, time consuming and environmentally intrusive (Mohamed *et al.*, 2018). Therefore, assessing and modelling the spatial distribution of SOC and identifying its key drivers is essential for developing short mid and long term management plans that improve soil health and ecosystem resilience in the country.

Conducting studies on SOC, complemented by the development of robust digital soil maps, is a critical necessity. Such studies provide the scientific basis needed to guide policy formulation, optimize resource allocation and ensure that interventions contribute meaningfully to climate change mitigation and sustainable agricultural livelihoods (Tesfay *et al.*, 2026). The present study addresses the above discussed knowledge gaps by developing a regional SOC prediction model to inform climate resilient land management in the Eritrean Highlands and Midlands. Focusing on an altitudinal gradient from the Central Highlands to the Western Midlands, we utilized soil, land use, geological, climatic, topographic and remote sensing data and compared three machine learning models: Cubist, random forest (RF) and partial least squares (PLS). These were selected for their demonstrated capacity to handle complex environmental datasets, allowing us to compare and provide alternative models.

These models have been successfully applied to predict SOC and have achieved promising results. For instance, Meliho *et al.* (2023) found that random forest

($R^2 = 0.79$, RMSE= 1.2) outperformed Cubist, SVM and Gradient Boosting models. In contrast, Suleymanov *et al.* (2023) reported Cubist as the most precise model for predicting soil organic matter ($R^2 = 0.64$, RMSE= 1.95). Similarly, Devine *et al.* (2020) identified multiple linear regression as a top performing algorithm in their studies. Collectively, these findings underscore the importance of testing multiple models, as such comparisons allow researchers to identify the most suitable approach for their specific context.

MATERIALS AND METHODS

Study region

The study area is situated in the upper watersheds of the Anseba and Barka rivers, Eritrea. It extends along the Asmara-Keren road, from Asmara (average altitude 2325 m a.s.l.) to Hamelmalo (128 m a.s.l.). Fig 1 displays the study area's location, elevation range, names of subzones and soil samples locations.

Mean annual rainfall ranges from around 400 to 500 mm and mean monthly temperature varies between 16.8°C and 24.6°C across the study area. Temperature exhibits a clear inverse relationship with altitude. Rainfall, however, follows a different spatial pattern, with the highest amounts recorded in the central parts of the watershed.

Mixed subsistence farming forms the primary livelihood in rural areas, supplemented by limited irrigated cropping. In the cooler upper reaches of the study area, highland crops such as barley, wheat and potato are common, while lower elevation sub zones (Elabered, Keren and Hamelmalo) predominantly cultivate lowland crops including sorghum, pearl millet and groundnut. Soils in the region are subject to heavy grazing and remain largely bare for much of the year.

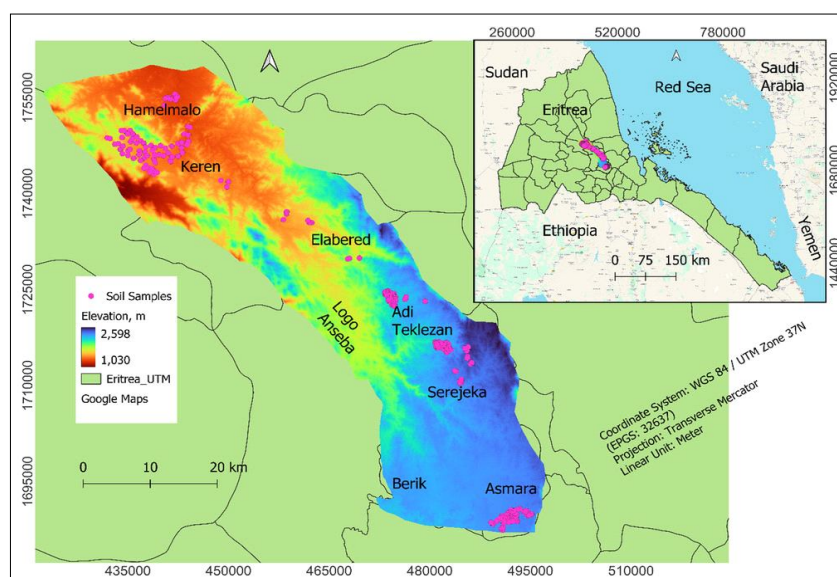


Fig 1: Study area and soil samples locations, subzones' names and elevation.

Soil, land use, soil taxonomic and geological data

The soil data were compiled from multiple sources: (1) newly collected samples from this study ($n = 113$), (2) legacy data obtained from the National Agricultural Research Institute (NARI) soil laboratory and (3) previously published datasets from studies conducted in parts of the target area (Tesfay *et al.*, 2020, $n = 21$; Nuguse *et al.*, 2019, $n = 12$).

From August to September 2023, surface soil samples (0-30 cm) were collected from the Adi Teklezan ($n = 49$) and Keren ($n = 64$) sub zones using a stratified sampling design (Tesfay *et al.*, 2024). Sampling spanned multiple land uses and topographic positions. Detailed protocols for soil sampling and preparation are described in Tesfay *et al.* (2025b). Soil organic carbon was analysed using the Walkley Black method (FAO, 2019) at the soil laboratory of the National Agricultural Research Institute (NARI) and particle size distribution (hydrometer method), pH (pH meter) and electrical conductivity (EC meter) and bulk density were determined at the soil laboratory of Hamelmalo Agricultural College.

The legacy soil data from NARI were merged with the other datasets and checked for completeness. Outliers in SOC values were identified using Z score statistics (Tesfay *et al.*, 2025a) and observations beyond ± 2 standard deviations were excluded. The final compiled dataset consisted of 245 georeferenced sampling points, each with recorded SOC, pH, sand, clay, silt, texture and electrical conductivity (EC) values.

Soils in the study area were classified as Cambisols, Lixisols, Leptosols and Fluvisols based on the harmonized world soil database (HWSD) (Tesfay *et al.*, 2025a). Soil taxonomic units are commonly employed for SOC prediction (Ayala Izurieta *et al.*, 2021; Tesfay *et al.*, 2025a).

Geological data were extracted from the Geological Map of Eritrea (Tesfay *et al.*, 2025a). Three parent material types were identified; intrusive sediments, metavolcanic and sedimentary rocks. Similarly geological units are commonly used in SOC prediction studies (Ayala *et al.*, 2021; Tesfay *et al.*, 2025a).

Land use history was recorded during soil sampling for newly collected samples. For legacy soil data lacking such records, land use categories were assigned using field knowledge of the study area in combination with Google Earth Pro imagery (Tesfay *et al.*, 2025a). Four land use types were represented: Enclosures (EC), Irrigated Farming (IF), rainfed farming (RF) and communal grazing (CG).

Climatic and topographic data

Long term climate records are scarce in Eritrea. Therefore, temperature and rainfall data were obtained from WorldClim 2.1 at 30 arc second resolution (Fick and Hijmans, 2017). Climate variables are widely recognized as important predictors in SOC modelling (Shen *et al.*, 2023; Galluzzi *et al.*, 2024; von Fromm *et al.*, 2024).

The study area exhibits strong elevational gradients and rugged terrain. Accordingly, topographic variables including altitude, slope and terrain roughness index along

others were derived from a 30 m resolution SRTM DEM. Such topographic metrics are commonly used in SOC prediction studies (Zhou *et al.*, 2023; Jendoubi *et al.*, 2025).

Remote sensing data

Two Landsat 8 scenes were obtained from the USGS archives <https://earthexplorer.usgs.gov/> with path/row and acquisition dates of 169/049 (09 March 2024) and 170/049 (17 April 2024). Both images had <10% cloud cover. Several spectral indices were derived, including the Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI) and Bare Soil Index (BSI) among others. In addition, Landsat 8 bands: Short Wave Infrared 1 and 2 (SWIR1, SWIR2), Near Infrared (NIR), Red, Green, Blue and bands 1 and 10 were included. Different studies have employed spectral data for SOC prediction along with other environmental covariates (Zhang *et al.*, 2023; Yami and Swami, 2023; Tesfay *et al.*, 2025a). For more information about the input variables please refer Tesfay *et al.* (2025a).

Selection of machine learning models and input variables

Machine learning algorithms such as random forest (RF), multiple linear regression (MLR), Cubist and partial least squares (PLS) are widely used in SOC prediction. Since relationships between SOC and predictor variables can be linear, non linear, or a combination of both, we selected one linear based model (PLS) and two decision tree based models (Cubist and RF) to capture these different functional forms.

Including a comprehensive set of potential predictors is important, but using too many variables can introduce multicollinearity, reduce model interpretability and lead to unreliable predictions. Therefore, reducing dimensionality while retaining informative predictors is essential. Our dataset initially contained 46 predictor variables and one target (SOC). Thus, we applied three feature selection approaches:

- 1. No selection:** All 46 proposed variables were used.
- 2. Correlation based filtering:** Variables with very high collinearity ($r > 0.900$) or no significant correlation with SOC ($p > 0.05$) were removed (Hosseinpour-Zarnaq *et al.*, 2024; Tesfay *et al.*, 2025a), retaining 28 variables.
- 3. Boruta algorithm feature selection:** This wrapper method selected 11 variables: B10, temperature, SOC1, land use (LU), sand, pH, BR2, rainfall, CI, altitude and negative openness (NvOp).

Model training, testing and evaluation

Fig 2 presents a comprehensive workflow of the methodology employed in this study. All modelling was conducted using Python 3.11.7. The dataset was randomly split into training (80%) and testing (20%) subsets. Three machine learning algorithms were applied: Partial least squares, cubist and random forest. This multi model approach enabled performance comparison and assessment of predictor importance.

Model performance was evaluated using the root mean square error (RMSE), coefficient of determination (R^2), ratio

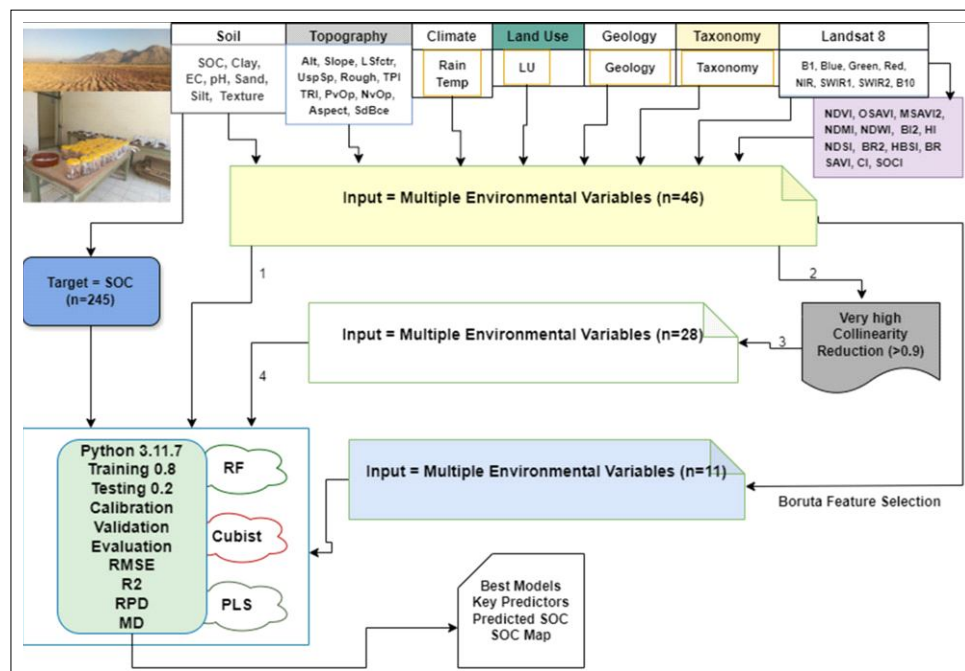


Fig 2: Conceptual flowchart of the methodologies used.

of performance to deviation (RPD (Chen *et al.*, 2024) and model discrepancy (MD) (Tesfay *et al.*, 2025a), as defined in equations 1-4 below. Hyperparameters were optimized via five-fold GridSearchCV.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad \dots(1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad \dots(2)$$

$$RPD = \frac{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}}{\sqrt{\sum_{i=1}^n (O_i - P_i)^2}} \quad \dots(3)$$

$$MD = \frac{T - t}{T} * 100\% \quad \dots(4)$$

Where,

n = The count for observed SOC.

O_i and P_i = The observed and predicted SOC values, respectively.

\bar{O} = The arithmetic mean for observed SOC.

T and t = The RMSE values for the training and testing datasets, respectively.

RPD indicates the extent to which the capacity of predictions improves in comparison to simply using the mean of the original dataset.

RESULTS AND DISCUSSION

Descriptive of soil parameters

The soils contained very high sand (mean = 63.56%) and low silt (23.93%) and clay (12.58%) contents (Table 1). Clay and silt exhibited high spatial variability with coefficients of variation (CV) of 78.49% and 52.91%, respectively. The predominance of coarse textures can be partly attributed to the region's steep, rugged topography and sparse vegetation, which promote downslope transport of runoff and fine soil particles.

Mean electrical conductivity (EC) and pH were 0.12 dS m⁻¹ and 7.41, respectively. These indicate non saline and mildly alkaline soil conditions, which are conducive for most plants and crops.

The soils exhibited mean SOC of 0.71%, ranging from 0.04% to 1.53% (SD = 0.37, CV = 52.48%). SOC showed strong positive correlation with altitude ($r = 0.580$)

Table 1: Descriptive of observed soil parameters.

Parameter	Minimum	Maximum	Mean	SD	CV, %
SOC, %	0.04	1.53	0.71	0.37	52.48
Sand, %	18.80	92.50	63.56	17.87	28.11
Silt, %	2.50	54.20	23.93	12.66	52.91
Clay, %	2.50	51.60	12.58	9.87	78.49
EC, dS m ⁻¹	0.02	0.35	0.12	0.06	51.61
pH	5.86	9.02	7.41	0.54	7.23

and negative with temperature ($r = -0.582$), which highlights the role of cooler, higher elevation environments in promoting soil carbon sequestration. This aligns with previous findings from the region (Nuguse *et al.*, 2019) and with global studies emphasizing the influence of temperature and elevation on SOC (Malla *et al.*, 2022; Chen *et al.*, 2023; Pellikka *et al.*, 2023). Similar poor SOC levels have been reported in the region (Nuguse *et al.*, 2019; Tesfay *et al.*, 2024) and in the Gash Barka region of the country (Tefay *et al.*, 2025a,b). These low levels likely result from a combination of factors including widespread unsustainable management practices such as crop residue removal, overgrazing, conventional tillage and continuous monocropping coupled with high erosion rates and an arid climate that limits soil development (Nuguse *et al.*, 2019; Tesfay *et al.*, 2024, 2025a,b).

Soil quality can be enhanced through crop residue retention, manure application, crop rotation, agroforestry, cover cropping, minimal tillage and biochar. For grazing lands, beneficial strategies include adequate resting, rotational grazing, enclosures, cut and carry systems and SOC dynamics models (Tefay *et al.*, 2024). In arid regions, external organic inputs via composting or biochar production are essential to improve soil health and productivity.

Effect of land use types on SOC

Different land use types exhibited significantly varied mean SOC levels. Soils from irrigated farming (IF) had the highest average SOC, 1.03%, followed by enclosures (EC), 0.84%, rainfed farming (RF), 0.59% and communal grazing (CG)

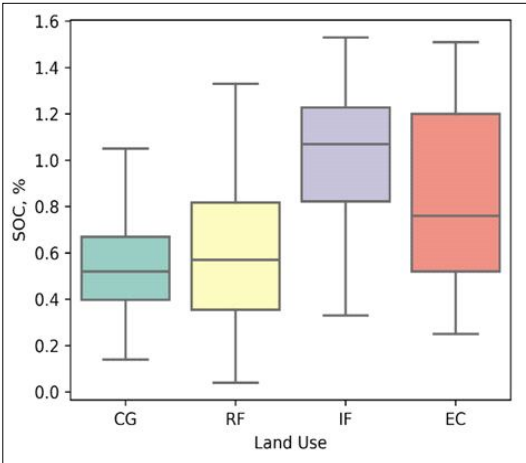


Fig 3: ANOVA analysis; land uses versus mean SOC levels.

lands, 0.57% (Fig 3). Tesfay *et al.* (2024) also found that SOC was notably higher in irrigated soils at Keren subzone, which is part of the study area. This indicates the necessity of irrigation and soil moisture for soil development and carbon sequestration. Irrigated farming systems benefit from superior agronomic management compared to rainfed farming, which leads to higher crop yields, enhanced carbon sequestration, reduced CO₂ emissions and decreased sensible heat (Weldewahid *et al.*, 2023; Abdellatif *et al.*, 2023). However, (Tefay *et al.*, 2025a) reported that irrigated farms had statistically as low SOC as that of RF and CG lands in the Western Lowlands of the country. Moreover, Haile *et al.* (2025) reported that even though irrigation expanded but yield couldn't at Dge subzone in the Western Lowlands but increased at Ghala Nefhi in the Highlands. Enclosures also demonstrated significant potential for carbon sequestration. Nuguse *et al.* (2019) and Tesfay *et al.* (2025a, 2026) also reported considerably higher SOC levels in natural forestlands and enclosures. However, Hoang (2024) reported that SOC levels were higher in cropping fields than in natural forestlands. These highlight the importance of management in all types of land use systems.

RF and CG soils displayed significantly lower SOC levels, 0.59% and 0.57%, respectively. This suggests their potential for carbon release, contributing to increased atmospheric CO₂ concentrations and global warming, which support the reports of various research (Nuguse *et al.*, 2019; Weldewahid *et al.*, 2023; Abdellatif *et al.*, 2023; Prabakaran *et al.*, 2023; Urgessa *et al.*, 2023; Tesfay *et al.*, 2024, 2025a,b, 2026). Therefore, converting rainfed to irrigated farming and grazing lands to enclosures, would provide environmental and socio-economic benefits. Additionally, endorsing enhanced management practices for IF and EC land uses is essential to maximize their capacity for carbon storage.

Performance of models

Cubist model with 46 variables (Cubist-46) attained the uppermost accuracy ($R^2 = 0.7465$, $RMSE = 0.1790$, $RPD = 2.0895$ and $MD = 0.21\%$). In contrast, PLS-11 model produced the lowest prediction accuracy ($R^2 = 0.5930$, $RMSE$ of 0.2368, $RPD = 1.6491$ and $MD = 11.53\%$) (Table 2, Fig 4).

Generally, the RF model showed the least model discrepancy (MD) changes when the number of input variables changed. This indicates its stability. The Cubist model exhibited a decreasing trend in MD when number of covariates decreased (Fig 4b) with MD values of 5.42, 4.02,

Table 2: Models' performance with different inputs for the validation data.

Model	46 variable		28 variables		11 variables	
	R ²	RPD	R ²	RPD	R ²	RPD
RF	0.6869	1.8803	0.7015	1.9257	0.6954	1.9063
Cubist	0.7465	2.0895	0.6992	1.9181	0.6703	1.8322
PLS	0.6101	1.6847	0.6364	1.7446	0.5930	1.6491

2.51% for the Cubist-46, Cubist-28 and Cubist-11 models, respectively. However, the lowest and highest MD were recorded by the PLS-46 and PLS-11 models with MD values of 0.21 and 11.53%, respectively. This indicates the model's sensitivity with changing number of covariates. The PLS model demonstrated a consistent decline in performance as the number of input variables decreased (Fig 4b). Model discrepancy (MD) represents the disparity in performance of a machine learning model when evaluated on testing data compared to training data. This metric is essential for determining how effectively a model captures unobserved data. The tolerance level for model discrepancy varies with the level of precision needed by the specific task (Tesfay *et al.* 2025a).

Based on the RPD scheme classification (Viscarra *et al.*, 2007), all the RF and Cubist models were categorized as good, whereas PLS models as fair. Consequently, the Cubist and RF models are suitable for SOC prediction, quantification and monitoring applications in the study region. The PLS models, while less accurate, may still be useful for broader, general purpose SOC assessments.

Variables' importance

Variable importance for SOC prediction, based on Random Forest permutation, is shown in Fig 5. When using 46 input variables, the ten most important predictors were (in order) temperature, land use, altitude, SOCI, B10, rainfall, BR2, positive openness (PvOp), SAVI and pH. Together, these accounted for 73.58% of the variance in SOC, with the top five explaining 57.41% and temperature alone contributing 19.73% (Fig 5a).

With 28 variables, the top ten predictors were temperature, land use, SOCI, B10, rainfall, BR2, PvOp, sand, MSAVI2 and NDVI, explaining 79.95% of SOC variance. Here, temperature contributed 26.67% and the top five variables explained 62.15% (Fig 5b).

Using the 11 variable set, the most important predictors were temperature, land use, SOCI, altitude, B10, BR2, rainfall, sand, negative openness (NvOp) and CI.

These collectively explained 97.01% of the variance, with the top five accounting for 71.96% and temperature alone representing 21.34% (Fig 5c).

Across all variable sets, temperature emerged as the most important predictor of SOC in the study area, which underscores the strong influence of climate on soil carbon dynamics in arid regions. This finding aligns with other studies worldwide that have identified temperature as a critical factor in SOC modeling (e.g. Galluzzi *et al.*, 2024; von Fromm *et al.*, 2024).

Fig 5 (d and e) illustrates variable importance by variable type category. Climate variables (temperature and rainfall) were the most influential, accounting for 25.08% of the explained variance. This was followed by topography (19.66%), bare-soil related indices (19.32%), land use (12.80%) and Landsat 8 bands (9.42%), among other categories. Temperature alone contributed 19.73%, while its surrogate variables altitude (11.10%), B10 (5.62%) and positive openness (3.27%) together raised the total contribution of temperature related predictors to 39.27%.

The same predictor variables were used in a previous study conducted in the Western Lowlands of Eritrea (Tesfay *et al.*, 2025a). In that setting, rainfall emerged as the most important predictor of SOC, whereas temperature dominated in the current highland study area. These findings underscore the need for climate smart agricultural planning that accounts for regional climatic differences to enhance ecosystem and societal resilience.

In the Western Lowlands, the key SOC predictors were (in order) rainfall, temperature, altitude, soil taxonomy, SWIR2, sand, clay and NIR (Tesfay *et al.*, 2025a). In the Highlands, the main predictors were temperature, land use, altitude, SOCI, B10, rainfall, BR2 and positive openness. Climate related variables (temperature, rainfall, altitude) were important in both regions, but land use ranked highly only in the Highlands. This likely reflects the longer history of established irrigated farming in the highlands, whereas irrigation in the lowlands is more recent and often poorly managed (Tesfay *et al.*, 2025a).

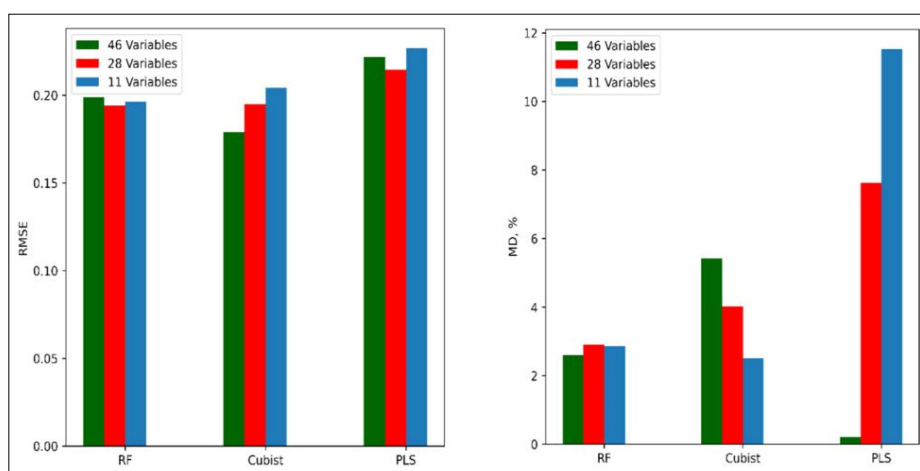


Fig 4: RMSE (left) for the testing data and MD (right) versus models utilizing different input variables.

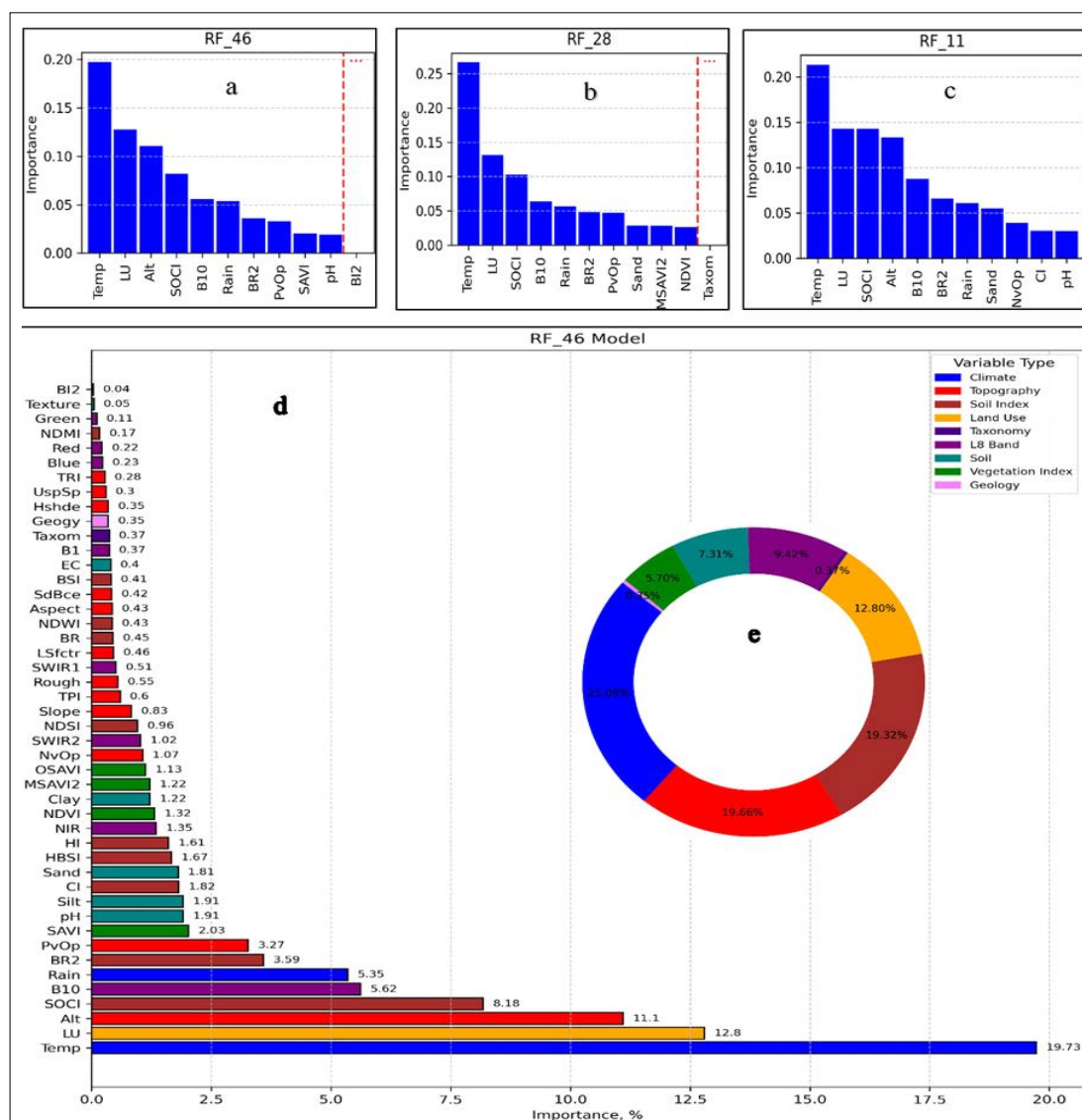


Fig 5: Importance of variables; RF model with 46 (a), 28 (b) and 11 (c) input variables, importance by variable type (d) and summarized by the pie chart (e).

Predicted SOC and mapping

The average predicted SOC by the Cubist-46 model was recorded at 0.70%, while the other two models had a mean of 0.71%. Related to the measured soil C, the skewness, CV and SD values for the predicted soil C decreased across all three models. Among them, the RF-28 model demonstrated the lowest CV (39.23%), followed by the PLS-28 model (42.13%) in contrast to the observed SOC with CV of 52.48%. Regression models tend to compress the data closer to the average.

Following SOC prediction using the Cubist-46 model, we applied regression kriging (RK) to generate spatial maps across the study area (Fig 6). SOC increased with

elevation (Fig 6a), a pattern inversely related to the temperature gradient across the study area. The lowest SOC levels were found in the lower elevation sub zones of Keren and Hamelmalo, whereas the highest concentrations occurred in the upper parts of the watershed. The highlands receive 450-500 mm annual rainfall with cooler temperatures that favour vegetation growth and reduce carbon decomposition, whereas the midlands receive 350-400 mm, with high temperature and evapotranspiration where aridity limits plant cover and accelerates SOC loss.

The RK approach effectively combined Cubist predictions with residual (observed SOC-predicted SOC)

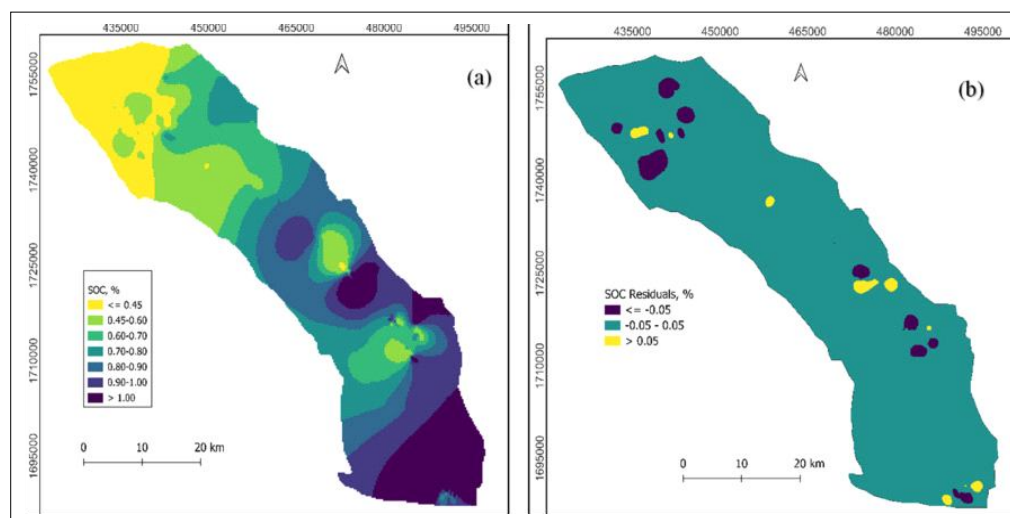


Fig 6: Predicted SOC (a) and SOC prediction residuals (b) spatial distribution produced through regression kriging following SOC prediction with the highest accuracy Cubist-46 model.

spatial interpolation. Residuals narrowed substantially from -0.41 to 0.44% before kriging to -0.08 to 0.07% afterward, with most values falling between -0.05 and 0.05% (Fig 6b). This reduction confirms that kriging successfully removed spatially structured errors, enhancing prediction accuracy.

CONCLUSION

The studied soils exhibit poor SOC levels (mean= 0.71%) with high spatial variability, underscoring the urgent need for remedial interventions. Land use exerted a significant influence on SOC ($p < 0.001$); irrigated farmlands and enclosures displayed significantly higher SOC than grazing and rainfed lands, highlighting the potential of well managed land use systems for soil carbon storage.

Among the evaluated models, the Cubist model with 46 variables ($R^2 = 0.7465$, $RPD = 2.0895$) outperformed both RF and PLS. The most important predictors of SOC were temperature, land use, altitude, SOCI, B10 and rainfall, demonstrating the dominant role of climate in shaping the spatial variation of SOC. This finding is supported by the strong negative correlation of SOC with temperature ($r = -0.582$) and positive correlation with altitude ($r = 0.580$).

In conclusion, this study demonstrates that in low capacity regions, SOC assessment and monitoring can be achieved at minimal cost using machine learning with remote sensing, climate and topographical variables. The developed Cubist and RF models can be effectively deployed to guide planning and monitoring of strategies aimed at improving soil fertility and food production, restoring degraded land and mitigating climate change impacts across the Central Highlands to the Western Midlands of Eritrea.

ACKNOWLEDGEMENT

The study was supported by grant No. 25-46-02010 from the Russian Science Foundation, <https://rscf.ru/project/25-46-02010/>.

Disclaimers

The views and conclusions expressed in this article are solely those of the authors and do not necessarily represent the views of their affiliated institutions.

Conflict of Interest

The authors declare no conflicts of interest.

REFERENCES

- Abdellatif, M.A., Hassan, F.O., Rashed, H.S.A., El Baroudy, A.A., Mohamed, E.S. *et al.* (2023). Assessing soil organic carbon pool for potential climate-change mitigation in agricultural soils: A case study fayoum depression, Egypt. *Land*. **12**(9): 1755. <https://doi.org/10.3390/land12091755>.
- Ayala, I.J.E., Márquez, C.O., García, V.J., Jara, S.C.A., Sisti, J.M. *et al.* (2021). Multi-predictor mapping of soil organic carbon in the alpine tundra: A case study for the central Ecuadorian páramo. *Carbon Balance and Management*. **16**(1): 1-19. <https://doi.org/10.1186/s13021-021-00195-2>.
- Chen, D., Yu, M., González, G. and Gao, Q. (2023). Altitudinal Pattern of Soil Organic Carbon and Nutrients in a Tropical Forest in Puerto Rico. In *Neotropical Gradients and their Analysis*. https://doi.org/10.1007/978-3-031-22848-3_12.
- Chen, Q., Wang, Y. and Zhu, X. (2024). Soil organic carbon estimation using remote sensing data-driven machine learning. *Peer J*. **12**(8). <https://doi.org/10.7717/peerj.17836>.

- Devine, S.M., O'Geen, A.T., Liu, H., Jin, Y., Dahlke, H.E., Larsen, R.E. and Dahlgren, R.A. (2020). Terrain attributes and forage productivity predict catchment-scale soil organic carbon stocks. *Geoderma*. **368**. <https://doi.org/10.1016/j.geoderma.2020.114286>.
- FAO. (2021). National Agricultural Innovation System Assessment in Eritrea-Consolidated Report. Rome, Italy <https://doi.org/10.4060/cb7296en>.
- FAO. (2017). Soil Organic Carbon: The Hidden Potential (A.V.B.R.V.R. Wiese Liesl, Ed.). Food and Agriculture Organization of the United Nations, Rome, Italy.
- FAO. (2019). Standard Operating Procedure for Soil Organic Carbon: Walkley-Black method: Titration and Colorimetric Method (1st ed.). Global Soil Laboratory Network GLOSOLAN, Rome, Italy, 2019.
- Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. **37(12)**: 4302-4315. <https://doi.org/10.1002/joc.5086>.
- Galluzzi, G., Plaza, C., Priori, S., Giannetta, B. and Zacccone, C. (2024). Soil organic matter dynamics and stability: Climate vs. time. *Science of The Total Environment*. **929**: 172441. <https://doi.org/10.1016/j.scitotenv.2024.172441>.
- Ghebregabher, M.G., Yang, T., Yang, X. and Wang, C. (2019). Assessment of desertification in Eritrea: Land degradation based on landsat images. *Journal of Arid Land*. **11(3)**. <https://doi.org/10.1007/s40333-019-0096-4>.
- Haile, B.T., Ramoelo, A., Dougill, A.J. and Qabaqaba, M. (2025). Land use/land cover (LULC) change and irrigated area monitoring in eritrea: Insights into horticultural production and sustainability. *Remote Sensing in Earth Systems Sciences*. <https://doi.org/10.1007/s41976-025-00247-y>.
- Hoang, T.V.H. (2024). Land use practices effects on soil organic carbon and nitrogen in North-central Vietnam. *Agricultural Science Digest*. doi: 10.18805/ag.DF-623.
- Hosseinpour-Zarnaq, M., Moshiri, F., Jamshidi, M., Taghizadeh-Mehrjardi, R., Tehrani, M. *et al.* (2024). Monitoring changes in soil organic carbon using satellite-based variables and machine learning algorithms in arid and semi-arid regions. *Environmental Earth Sciences*. **83(20)**. <https://doi.org/10.1007/s12665-024-11876-9>.
- Jendoubi, D., Liniger, H. and Speranza, C.I. (2025). Impacts of Land Use and Topography on Soil Organic Carbon in a Mediterranean Landscape (North-Western Tunisia). <https://doi.org/http://dx.doi.org/10.5194/soil-5-239-2019>.
- Malla, R., Neupane, P.R. and Köhl, M. (2022). Modelling soil organic carbon as a function of topography and stand variables. *Forests*. **13(9)**. <https://doi.org/10.3390/f13091391>.
- Meliho, M., Boulmane, M., Khattabi, A., Dansou, C.E., Orlando, C.A. *et al.* (2023). Spatial prediction of soil organic carbon stock in the moroccan high atlas using machine learning. *Remote Sensing*. **15(10)**: 2494. <https://doi.org/10.3390/rs15102494>.
- MoA. (2018). National Land Degradation Neutrality Targets, Ministry of Agriculture, (MoA) Asmara, Eritrea.
- Mohamed, E.S., Saleh, A.M., Belal, A.B. and Gad, A.A. (2018). Application of Near-infrared reflectance for quantitative assessment of soil properties. *Egyptian Journal of Remote Sensing and Space Science*. **21(1)**. <https://doi.org/10.1016/j.ejrs.2017.02.001>.
- Nuguse, M.T., Singh, B. and Ogbazghi, W. (2019). Studies on soil organic carbon and some physico-chemical properties as affected by different land uses in Eritrea. *Journal of Soil and Water Conservation*. **18(3)**: 213. <https://doi.org/10.5958/2455-7145.2019.00030.4>.
- Page, K.L., Dang, Y.P. and Dalal, R.C. (2020). The ability of conservation agriculture to conserve soil organic carbon and the subsequent impact on soil physical, chemical and biological properties and yield. *In Frontiers in Sustainable Food Systems*. **4**. <https://doi.org/10.3389/fsufs.2020.00031>.
- Pellikka, P., Luotamo, M., Sädekoski, N., Hietanen, J., Vuorinne, I. *et al.* (2023). Tropical altitudinal gradient soil organic carbon and nitrogen estimation using Specim IQ portable imaging spectrometer. *Science of The Total Environment*. **883**: 163677. <https://doi.org/10.1016/j.scitotenv.2023.163677>.
- Prabakaran, S., Kaleeswari, R.K., Backiyavathy, M.R., Jagadeeswaran, R., Selvi, R.G. and Bama, K.S. (2023). Soil carbon stock and carbon pool indices under major land use systems of Mayiladuthurai District, Cauvery Delta Zone, India. *Agricultural Science Digest*. doi: 10.18805/ag.D-5786.
- Shen, C., Xiao, W., Chen, J., Hua, L. and Huang, Z. (2023). Climate-sensitive spatial variability of soil organic carbon in multiple forests, Central China. *Global Ecology and Conservation*. **46**: e02555. <https://doi.org/10.1016/j.gecco.2023.e02555>.
- Suleymanov, A., Tuktarova, I., Belan, L., Suleymanov, R., Gabbasova, I. and Araslanova, L. (2023). Spatial prediction of soil properties using random forest, k-nearest neighbours and cubist approaches in the foothills of the Ural Mountains, Russia. *Modeling Earth Systems and Environment*. **9(4)**: 3461-3471. <https://doi.org/10.1007/s40808-023-01723-4>.
- Tesfay, T., Mohamed, E.S., Mehrteab, M., Ghebrensaie, T.W. and Sereke, T.E. (2025b). Soil organic carbon losses following conversion of natural forests into agriculture: Insights from Eritrea. *Dokuchaev Soil Bulletin*. **123**: 100-115. <https://doi.org/10.19047/0136-1694-2025-123-100-115>.
- Tesfay, T., Mohamed, E.S., Savin, I.Y., Kucher, D.E., Rebouh, N.Y. and Ogbazghi, W. (2025a). Soil organic carbon modelling with different input variables: The case of the western lowlands of eritrea. *Sustainability*. **17(21)**: 9884. <https://doi.org/10.3390/su17219884>.
- Tesfay, T., Ogbazghi, W. and Singh, B. (2020). Effects of soil and water conservation interventions on some physico-chemical properties of soil in Hamelmalo and Serejeka Sub-zones of Eritrea. *Journal of Soil and Water Conservation*. **19(3)**: 229-234. <https://doi.org/10.5958/2455-7145.2020.00031.4>.
- Tesfay, T., Ogbazghi, W., Singh, B. and Tsegai, T. (2018). Factors influencing soil and water conservation adoption in basheri, gheshnashm and shmangus laelai, eritrea. *IRA-International Journal of Applied Sciences*. **12(2)**. <https://doi.org/10.21013/jas.v12.n2.p1>.
- Tesfay, T., Mohamed, E.S.S., Ghebrensaie, T.W., Ghebremariam, S.B. and Mehrteab, M. (2024). Soil organic carbon stock assessment for soil fertility improvement, ecosystem restoration and climate-change mitigation. *E3S Web of Conferences*. **555**. <https://doi.org/10.1051/e3sconf/202455501015>.

- Tesfay, T., Ghebretsaie, T. and Mohamed, E. (2026). Estimating carbon sequestration potential of dryland enclosures: A comparative assessment assisted by Sentinel 2 time-series and machine learning framework. *Eurasian J. Soil Sci.* **15(2)**: 209-227. <https://doi.org/10.18393/ejss.1881636>.
- Urgessa, H.T. and Ferede, T.G. (2023). Effect of land use on plant nutrient availability and soil carbon stock of mokonisa machi watershed, Dugda Dawa Woreda, West Guji Zone, Southern Ethiopia. *Agricultural Science Digest.* **43(2)**: 135-142. doi: 10.18805/ag.RF-232.
- Viscarra, R.R.A.V., Taylor, H.J. and McBratney, A.B. (2007). Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing. *European Journal of Soil Science.* **58(1)**. <https://doi.org/10.1111/j.1365-2389.2006.00859.x>.
- von Fromm, S.F., Doetterl, S., Butler, B.M., Aynekulu, E., Berhe, A.A. *et al.* (2024). Controls on timescales of soil organic carbon persistence across sub-Saharan Africa. *Global Change Biology.* **30(1)**: e17089. <https://doi.org/10.1111/gcb.17089>.
- Weldewahid, Y., Habtu, S., Taye, G., Teka, K. and Gessesse, T.A. (2023). Effects of long-term irrigation practice on soil quality, organic carbon and total nitrogen stocks in the drylands of Ethiopia. *Journal of Arid Environments.* **214**: 104982. <https://doi.org/10.1016/j.jaridenv.2023.104982>.
- Yami, B., Swami, S., Singh, N.J. and Handique, B.K. (2023). Mapping and Monitoring of Soil Organic Carbon using Regression Analysis of Spectral Indices. <https://www.currentscience.ac.in/Volumes/124/12/1431>.
- Zhang, S., Tian, J., Lu, X. and Tian, Q. (2023). Temporal and spatial dynamics distribution of organic carbon content of surface soil in coastal wetlands of Yancheng, China from 2000 to 2022 based on Landsat images. *Catena.* **223**. <https://doi.org/10.1016/j.catena.2023.106961>.
- Zhou, T., Lv, Y., Xie, B., Xu, L., Zhou, Y., Mei, T., Li, Y., Yuan, N. and Shi, Y. (2023). Topography and soil organic carbon in subtropical forests of China. *Forests.* **14(5)**: 1023. <https://doi.org/10.3390/f14051023>.