Pol. J. Environ. Stud. Vol. XX, No. X (XXXX), 1-12
DOI: 10.15244/pjoes/204570 ONLINE PUBLICATION DATE:

Original Research

Modeling Soil Carbon Variability along an Elevation Gradient in a Western Himalayan Mountain Ecosystem Using Multiple Statistical Approaches

Nouman Mubarak[1], Raja Waqar Ahmed Khan[1]*, Hamayun Shaheen[1], Tariq Saiff Ullah[2], Tanveer Hussain[3], Javeed Hussain[2], Muhammad Nasir[2], Shazia Khatoon[4], Muhammad Manzoor[5], Wayne Thomas Shier[6]**

[1]Department of Botany, The University of Azad Jammu and Kashmir, Muzaffarabad, 13101. Pakistan*
[2]Department of Botany, University of Kotli, Azad Jammu and Kashmir, 11100. Pakistan**
[3]Department of Botany, Mirpur University of Science and Technology (MUST), Mirpur 10250 (AJK), Pakistan
[4]Department of Botany, Women University of Azad Jammu and Kashmir, Bagh, 12500, Pakistan.
[5]Department of Plant Sciences, Quaid-i-Azam University, Islamabad, 45320, Pakistan
[6]Department of Medicinal Chemistry, College of Pharmacy, University of Minnesota, 308 Harvard St., SE, Minneapolis, MN 55455, USA.

Received: 2 December 2024
Accepted: 1 May 2025

Abstract

Understanding soil organic carbon dynamics is crucial for climate change mitigation and developing strategies for sustainable land management. This research aimed to quantify SOC, analyze the effect of physicochemical properties on SOC, and evaluate different statistical models to identify the effective predictors for SOC. A total of 280 soil samples were analyzed for physicochemical properties using standard protocols. The study applied several statistical models, including Linear Mixed, Random Forest, Bayesian Linear, Generalized Additive, Multivariate Regression models, and Principal Component Analysis (PCA). Undisturbed soils exhibited significantly higher SOC stocks, averaging 74.71±8.65 Mg ha^{-1}, compared to 53.58±7.13 Mg ha^{-1} in disturbed soils. The LMM and GAM indicated a significant baseline SOC but showed no notable effect of altitude on SOC ($p = 0.703$ and 0.62-0.93, respectively). RFR identified bulk density as the strongest predictor, with the highest node purity (2269.57). PCA accounted for 78.35% of the variance, showing the critical role of soil texture in stabilizing SOC. Altitude showed a minimal effect; soil bulk density is the key factor in SOC variability, while clay

*e-mail: rajawaqar345@gmail.com;
**e-mail: tariqbot89@gmail.com

content is crucial for SOC stabilization. Advanced models like RFR provide better SOC predictions, aiding sustainable land management while incorporating additional variables that could further enhance SOC forecasting.

Keywords: soil, carbon sequestration, Himalayas, bulk density, altitude, random forest model, climate change, sustainability

Introduction

Soil organic carbon (SOC) sequestration plays a key role in the global carbon cycle. Soil is the largest terrestrial carbon pool, containing approximately 2500 Pg of carbon, significantly more than the atmospheric carbon pool of about 760 Pg [1]. Increasing SOC stocks through practices such as improved land management can help sequester carbon dioxide (CO_2) from the atmosphere, thus mitigating climate change [2]. Healthy soils contribute to productivity, biodiversity, and water retention and are vital for food security and resilience to climate impacts [3].

Appropriate soil management practices can reduce carbon emissions from soils by up to 50% by 2050 compared to 2010 levels [1]. Despite the potential, implementing effective SOC management strategies can be challenging. These include the need for long-term commitment to soil management, the vulnerability of certain soils to carbon loss, and the complexity of SOC dynamics influenced by climate [4].

The Himalayan mountainous ecosystems are classified as a biodiversity hotspot, with varying climatic and topographical conditions [5]. The climatic diversity in this region results from its complex geography and altitude, which influence climate and habitat types. The diverse plant communities in the Himalayas contribute significantly to SOC stocks through photosynthesis and organic matter inputs to the soil [6]. Soils in these ecosystems can store substantial amounts of carbon to mitigate climate change by reducing atmospheric CO_2 levels [7].

Climate change poses threats to the Himalayan ecosystems, potentially reducing soil health and SOC storage capacity. Increased temperatures and altered precipitation patterns can affect the delicate balance of these ecosystems, making sustainable management practices essential to protect and enhance SOC stocks [3, 5]. Soil erosion significantly contributes to SOC loss, adversely affecting soil health and intensifying climate change. Erosion, primarily caused by water, wind, and tillage, removes topsoil and depletes soil organic matter (SOM), which is essential for carbon storage [8].

To mitigate the effects of soil degradation, implementing conservation practices such as no-till farming, cover cropping, and maintaining vegetation can help minimize SOC loss, enhance soil structure, and improve the capacity to sequester carbon, promoting healthier ecosystems [7]. Studies on SOC variability in the Himalayan mountainous ecosystems tend to generalize findings across large areas, neglecting the specific impact of types and physicochemical attributes of soil along with altitude [9]. Additionally, there is a reliance on traditional statistical methods that may not fully capture the complexity of interactions between environmental variables and a lack of comprehensive data on the influence of soil disturbances on changes in SOC [10, 11].

The complexity of SOC dynamics in mountainous regions necessitates the use of robust and multifaceted methodological approaches, as traditional linear regression models, while useful, often fail to capture the complex and non-linear relationships between SOC and its environmental drivers [12, 13]. Therefore, advanced statistical modeling techniques are essential to accurately analyze the complex interactions between altitudinal variations and soil physicochemical attributes. These models improve our understanding of how different factors influence SOC, leading to better predictions and more effective soil management strategies [14].

This research uses multivariate regression and advanced statistical methods to model SOC dynamics in a representative Himalayan mountainous ecosystem, integrating physicochemical variables and altitude to comprehensively understand SOC variation. It also identified key soil properties that impact SOC and investigated potential non-linear interactions between SOC and other variables. These findings are crucial for predicting SOC responses to environmental and land-use alterations, especially pertinent in climate-sensitive mountainous ecosystems.

Materials and Methods

Study Area

The present study was conducted in the Makra mountainous region, located at 34° 34.461’ N and 73° 29.749’ E, with an altitudinal range from 1100 to 3860 m above sea level. The study area is situated in the Muzaffarabad district of Azad Jammu and Kashmir (AJK) (Fig. 1). This Himalayan mountainous ecosystem features a varied climate across different altitudes. At lower elevations (1000-2500 m), the climate is subtropical to temperate with mild summers and chilly winters, supporting subtropical and temperate forests. Mid-altitudes (2500-3500 m) experience cooler temperatures and increased snowfall, leading to sub-alpine meadows and scattered coniferous trees. The climate is alpine at high altitudes (above 3500 m) with year-round cold temperatures and minimal vegetation adapted to extreme conditions [15, 16].

The region experiences an average annual rainfall of 1457 mm, with an average temperature of 20.2°C. The maximum temperature recorded is 45°C in the lower altitudes, while the minimum drops below 0°C, leading to snowfall in the mountains [17]. The region exhibits a variety of soil types, ranging from deep alluvial soils in valleys to thin, rocky soils at higher altitudes. The habitat heterogeneity and climatic variations support a wide range of plant species (from subtropical forests to alpine meadows). These variations also influence the capacity of soils to sequester carbon [18].

Data Collection

Sampling was conducted along an altitudinal gradient from 1160 m to 3860 m (Fig. 1). The sampling points' altitude, latitude, and longitude were measured using a GPS device. Soil samples were collected using a 1×1 m quadrat at 100 m altitude intervals. A metallic core sampler was used to obtain soil samples. Five samples (0-30 cm depth) representing different slope classes were collected from each quadrat. Additionally, 5 more samples were randomly taken from nearby disturbed and eroded areas, resulting in a total of 10 samples per sampling site. In total, 280 soil samples were extracted from the field (28 sampling sites × 10 samples per site). The soil samples from each quadrat were thoroughly mixed to create a homogeneous sample, which was stored in plastic zipper bags. These samples were then brought to the Plant Ecology and Environmental Science Laboratory of the Department of Botany, the University of Azad Jammu and Kashmir, for physicochemical analysis and SOC stock assessment.

Determination of Soil Physicochemical Properties

Soil samples were crumbled and mixed thoroughly after being left open for 24 hours. A 20 mg portion of soil was weighed, and 100 ml of distilled water was added. The mixture was shaken for 2 minutes and left to settle for 5 minutes. Soil pH and Electroconductivity (EC) were measured using a pH/EC meter. The electrode was dipped 1-2 cm into the sample solution, and the pH value was recorded. Soil texture was evaluated using a jar test to determine the percentages of sand, silt, and clay. The soil textural triangle was then used to classify the soil type [19], while the soil bulk density (BD) was calculated using the following formula [20]:

$$BD\ (g/cm3) = \frac{\text{Weight of oven-dried soil (g)}}{\text{Volume of core sampler (cm3)}}$$

SOC was estimated using the Walkley and Black [21] wet oxidation method, known for its rapid and cost-effective analysis.

The percentage of SOM was calculated using the formula [22]:

$$\text{SOM}(\%) = (1 - \text{S/B}) \times 10 \times 0.68$$

SOC was determined by dividing the SOM percentage by a conversion factor of 2 [23]:

$$\text{SOC}\ (\%) = \text{SOM}\ (\%) \times 2$$

SOC was then converted into total SOC stock (Mg ha^{-1}) using the following Equation [22]:

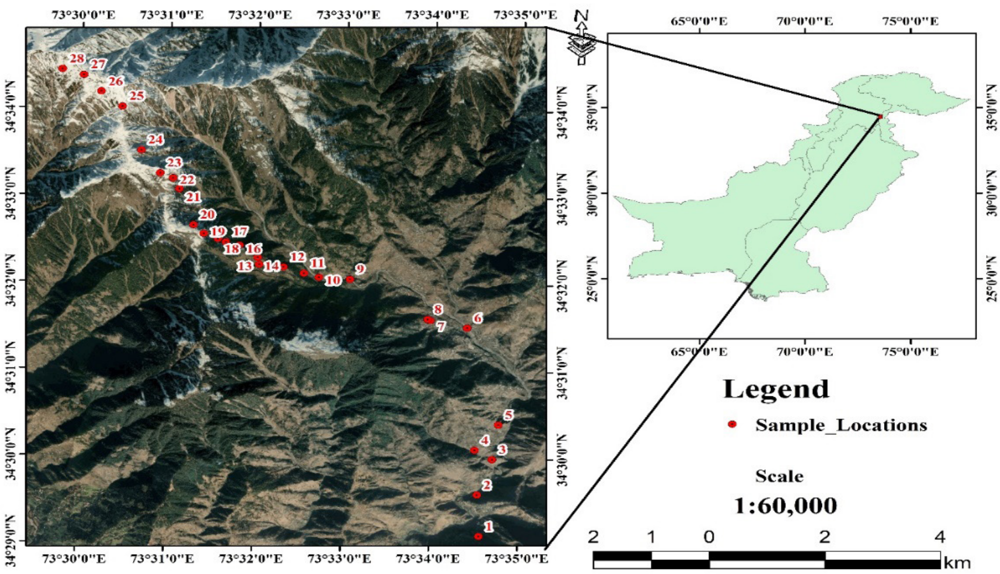

Fig. 1. Map of the study area indicating the sampling points along an elevation gradient.

$$SOC(Mg/ha) = Organic\ Carbon\ (\%)$$
$$\times\ Soil\ Bulk\ Density\ (g/cm3) \times Soil\ depth\ (cm)$$

Data Analysis and Modeling Techniques

All variables were standardized to ensure comparability across different measurement scales. Statistical techniques including Principal Component Analysis (PCA), Linear Mixed-Effects Model (LMM), Random Forest Regression (RFR), Bayesian Linear Modeling (BLM), Generalized Additive Model (GAM), and Multivariate Regression Analysis (MRA) were employed to explore the relationships between SOC and the various soil properties. Before applying the modeling techniques, multicollinearity was evaluated using Variance Inflation Factors, or VIF [24].

Each model was thoroughly validated using standard diagnostic techniques, including residual plots, R-squared values, and cross-validation. The significance of model coefficients was assessed, and model performance was compared across different statistical approaches to ensure reliability. All statistical analyses were conducted using R software version 4.4.0 [25]. To assess the performance of all models, a 10-fold cross-validation was employed. The dataset was divided into 10 subsets, with one subset used as the test set and the remaining nine as the training set in each iteration. This was repeated ten times, ensuring that every subset served as the test set once. The Root Mean Square Error (RMSE) was calculated for each fold's predictions, and the mean RMSE across all folds was used to gauge overall model accuracy. Consistent factor levels were maintained across the training and test datasets to ensure valid assessments [26, 27].

Results and Discussion

Soil Carbon Stock and Other Physicochemical Properties

The overall average SOM content was 3.54±0.63%, with values of 4.10±0.77% for undisturbed soil and 2.98±0.85% for disturbed soil. The average SOC content was 1.77±0.38%, higher in undisturbed soil at 2.05±0.46% compared to 1.49±0.42% in disturbed soil. The average SOC stock was 64.14±7.16 Mg ha^{-1}, with undisturbed soil containing 74.71±8.65 Mg ha^{-1} and disturbed soil 53.58±7.13 Mg ha^{-1}. The maximum SOC of 101.15±10.21 Mg ha^{-1} was recorded at 2960 m and a minimum of 34.99±4.27 Mg ha^{-1} at 2760 m (Table 1). Bivariate analysis showed that SOC did not significantly correlate with altitude in the study area (Fig. 2).

The mean soil pH was 6.39±0.8, with undisturbed soil at 6.36±0.74 and disturbed soil at 6.42±0.9. The highest pH (8.13±1.5) occurred at 1360 m, and the lowest (5.44±1.2) at 3880 m. The average soil EC was 20.24±4.62 µS cm^{-1}, with undisturbed soil showing 20.38±4.83 µS cm^{-1} and disturbed soil 20.10±4.41 µS cm^{-1}. EC peaked at 44.25±7.54 µS cm^{-1} at 1560 m and dropped to 7.13±2.21 µS cm^{-1} at 3660 m. The average soil BD was 1.20±0.6 g cm^{-3}, with undisturbed soil at 1.21±0.32 g cm^{-3} and disturbed soil at 1.19±0.15 g cm^{-3}. The maximum BD was recorded as 1.43±0.7 g cm^{-3} at 1660 m, and the minimum was 1.07±0.3 g cm^{-3} at 3660 m. (Table 1). In undisturbed soil, sandy loam was the most common, while loamy sand was the least common soil type. In disturbed soils, sandy loam was most frequently reported, and sand was least frequently recorded.

Principal Component Analysis (PCA)

PCA revealed that the first three components (PC1, PC2, and PC3) described most of the variance in the data, cumulatively capturing 78.35%, with standard deviations of 1.5864, 1.4455, and 1.2891, respectively. PC1 explains 31.5% of the variance, while PC2 and PC3 account for 26.1% and 20.7%, respectively. The fourth to sixth components (PC4, PC5, and PC6) contributed less, explaining 11.3%, 5.9%, and 3.03% of the variance, respectively, bringing the cumulative variance to 98.6%. PC7 and PC8 contributed negligibly, with the cumulative proportion reaching 100%. The PCA biplot illustrated the distribution of variables and samples across PC1 and PC2, which together explain a significant portion of the total variance (Fig. 3).

Table 1. Average soil carbon stock and physicochemical properties.

No.	Property	Undisturbed	Disturbed	Average
1	SOM %	4.10±0.77	2.98±0.85	3.54±0.63
2	SOC %	2.05±0.46	1.49±0.42	1.77±0.38
3	SOC Mg ha^{-1}	74.71±8.65	53.58±7.13	64.14±7.16
4	pH	6.36±0.74	6.42±0.9	6.39±0.8
5	EC µS cm^{-1}	20.38±4.83	20.10±4.41	20.24±4.62
6	BD g cm^{-3}	1.21±0.32	1.19±0.15	1.20±0.6

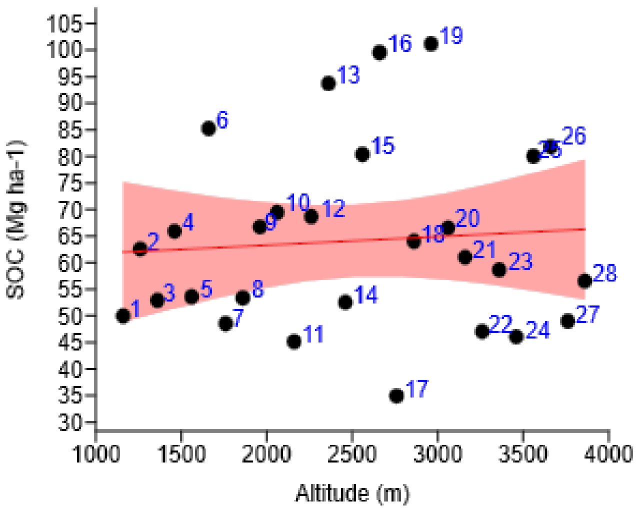

Fig. 2. Relationship between site-wise SOC values and altitude.

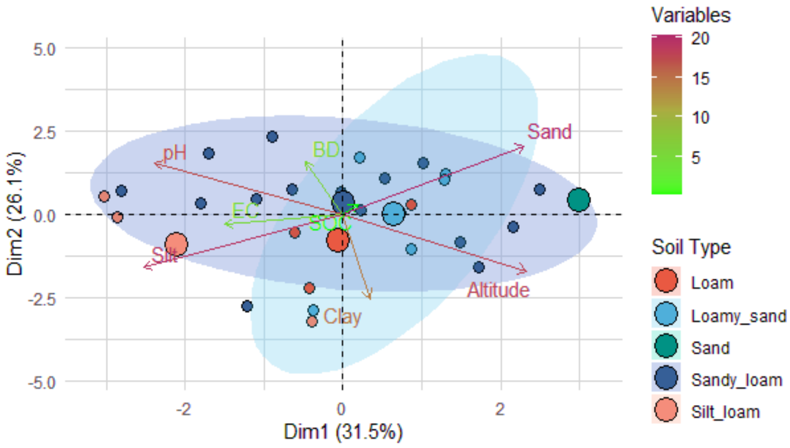

Fig. 3. PCA biplot depicting the relationships between soil variables, highlighting their distribution and contributions to overall variance.

Linear Mixed-Effects Model (LMM)

The LMM results revealed that the intercept was 60.21 (standard error = 10.74), with a t-value of 5.61 and a p-value of 6.81×10^{-6}, demonstrating a statistically significant baseline level of SOC. Conversely, the effect of altitude was estimated at 0.0016 (standard error = 0.0041), with a t-value of 0.39 and a p-value of 0.703, indicating that altitude does not have a statistically significant impact on SOC (Table 2). LMM suggested that while there is a significant baseline level of SOC, altitude does not meaningfully affect SOC within this model.

Generalized Additive Model (GAM)

The GAM results for SOC estimated the intercept at 64.78 (standard error = 11.02), with a t-value of 5.88 and a p-value of 6.44×10^{-6}. This model also showed a statistically significant baseline level of SOC. Loamy sand confirmed a decrease of -2.68 (standard error = 13.79), sand decreased by -11.01 (standard error = 22.04), and silt loam decreased by -3.33 (standard error = 16.25), while sandy loam increased by 1.04 (standard error = 12.02). However, in this model, none of the categorical variables (soil type) showed a statistically significant impact on SOC, with p-values ranging from 0.62 to 0.93 (Table 2). The smooth function of GAM for altitude captured non-linear relationships effectively.

Bayesian Linear Modeling (BLM)

The BLM estimated the intercept at 58.71, representing the SOC level at zero altitude with soil type as the reference category. The coefficient for altitude remained 0.0022, indicating a small positive effect on SOC; specifically, SOC increased by approximately 0.0022 Mg ha^{-1} for each meter increase in altitude; however, this effect is minor. The model also reveals significant variability in SOC based on soil type. Compared to the reference soil type, SOC decreased by 1.94 Mg ha^{-1} for loamy sand, 10.39 Mg ha^{-1} for sand, and

Table 2. Summary of Parameter Estimates for Linear Mixed (LMM) and Generalized Additive Models (GAM).

Model	Model Parameters	Estimate	Std. Error	t value	Pr(>|t|)
LMM	Intercept	60.21	10.7	5.6	6.81
	Altitude	0.002	0.004	0.4	0.7
GAM	Intercept	64.8	11	5.88	6.44
	Loamy sand	-2.7	13.8	-0.2	0.8
	Sand	-11	22	-0.5	0.6
	Sandy loam	1.03	12	0.09	0.9
	Silt loam	-3.3	16.2	-0.2	0.8

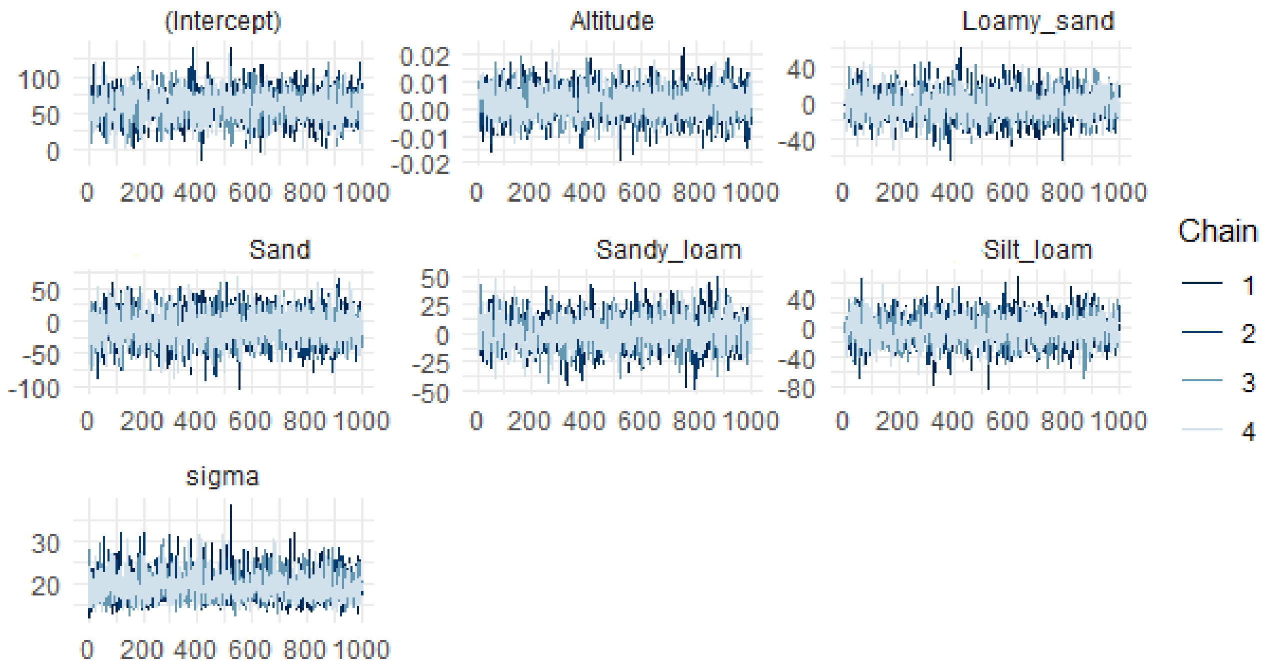

Fig. 4. Bayesian parameter estimates and uncertainty graph showing the variability and confidence intervals for the intercept, altitude, soil type, and error term (sigma).

2.85 Mg ha^{-1} for silt loam. Conversely, SOC increased by 1.76 Mg ha^{-1} for sandy loam. The BLM showed that soil type had a notable impact on SOC, with sand having the largest negative effect and sandy loam showing a moderate positive effect (Fig. 4).

Random Forest Regression (RFR)

The RFR model evaluated the relative importance of predictors on SOC using 1000 trees. The results indicated that BD is the most influential predictor, with an increased node purity of 2269.57, suggesting it had the strongest effect on SOC. Soil EC followed with an increase in node purity of 971.23, indicating a substantial impact on SOC. Soil pH showed an increase in node purity by 882.53, making it another significant predictor. Sand, silt, and clay content had increased in node purity by 662.60, 713.71, and 623.89, respectively, reflecting their moderate influence on SOC. Altitude showed the lowest increase in node purity at 524.30, suggesting it has a lesser impact on SOC than the other variables (Fig. 5).

Multivariate Regression Analysis (MRA)

The MRA was used to predict both SOC and pH simultaneously, utilizing altitude, soil type, BD, etc., as predictors. The model output showed how these predictors jointly influence SOC and soil pH. The predicted versus observed plots for both SOC and pH display an increasing trend, indicating a good alignment between the predictions of the model and the actual data. Additionally, the residual versus predicted values plots for SOC and pH exhibited a straight distribution

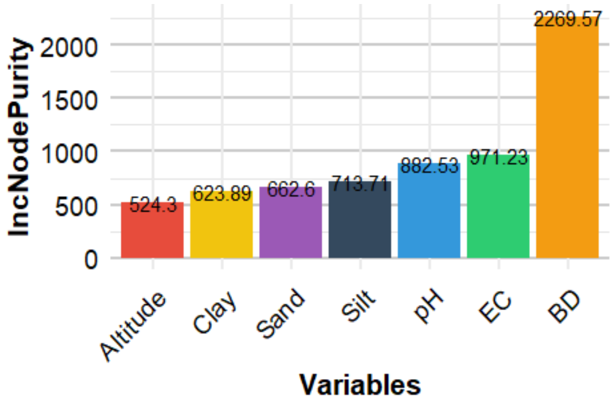

Fig. 5. Variable importance plot from the RFR model highlighting their relative importance to predict SOC.

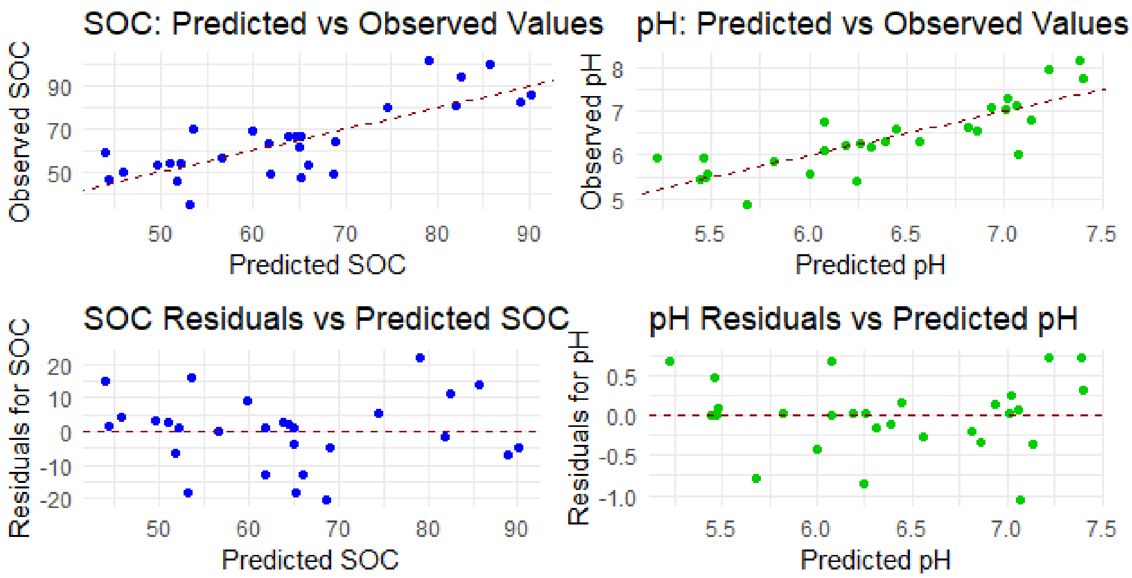

Fig. 6. Plots of the MRA showing predicted versus observed values and residuals versus predicted values for SOC and pH, demonstrating an even distribution of residuals around zero and indicating a well-fitting model.

around zero, indicating that the residuals are evenly spread and conform to the predictions. This straight-line pattern indicated that the model effectively captured the variance in the data and reflected a good fit with an appropriate error structure (Fig. 6).

SOC and Physicochemical Dynamics in Himalayan Mountain Ecosystems

SOC is a key component of the terrestrial carbon cycle, impacting soil health, carbon sequestration, and ecosystem stability [1]. SOC levels can differ significantly due to factors such as altitude, vegetation, climate, topography, and soil properties. Comparative research across various mountainous regions is crucial for understanding how these factors influence SOC dynamics [28]. The study recorded significant differences in soil physicochemical properties between undisturbed and disturbed areas, with higher SOM and SOC in undisturbed sites. Disturbances, topography, and heavy grazing led to SOC loss in disturbed areas [29]. Reduced SOC in degraded soils reflects carbon loss through oxidation, reduced organic inputs, and erosion-driven depletion [1, 30].

The literature review revealed significant spatial variability in SOC across the Himalayan region, influenced by elevation, vegetation, and land use. In the Indian Himalayas, SOC ranges between 50-150 Mg ha^{-1} [31], while in the Garhwal Himalayas, values between 124.8 and 185.6 Mg ha^{-1} have been observed [32]. In Himachal Pradesh, SOC varied from 36.04 to 73.26 Mg ha^{-1} [33], and in the broader Indian Himalayas, levels of 31.5 Mg ha^{-1} have been recorded [34]. In the Central Himalayas, SOC values between 63.9 and 83.8 Mg ha^{-1} are reported [35]. In the Punjab region of Pakistan, values of 30.19 Mg ha^{-1} [36] and in the Western Himalaya, SOC levels ranging from 27.18 to 34.89 Mg ha^{-1} have been found [37]. Other significant findings include soil carbon ranging from 261 to 370.6 Mg ha^{-1} in Kashmir, Western Himalaya [38], and from 56.7 to 337.8 Mg ha^{-1} in Bhutan Himalaya [39].

Lower SOC values are partly due to forest type, soil composition, and depth differences. In the study area, SOC loss is also driven by forest cover reduction and biomass extraction. In mountainous regions, steep slopes exacerbate soil degradation and impact SOC pools [40]. Forest composition and biomass production are influenced by physiographic and topographic factors, with elevation-induced climatic changes affecting precipitation patterns and soil characteristics [28]. Variations in soil structure also affect water retention and nutrient binding, leading to differences in SOC levels. Higher SOC levels enhance soil structure, stability, and fertility, reducing erosion and nutrient leaching [31]. Nonetheless, SOC is vulnerable to disturbances that can release CO_2 and deplete carbon stocks, indicating the importance of effective carbon sequestration and soil health management [3, 4]. SOC is higher in regions with favorable climatic and topographic conditions, dense forest cover, and minimal human disturbance [3]. Conversely, agriculture, deforestation, and land degradation reduce SOC, underscoring the need for sustainable land management practices to mitigate climate change [1].

Soil pH and EC remained stable between disturbed and undisturbed areas, with only minor increases in disturbed soils. Soil pH decreases with altitude and tends to be slightly acidic, promoting plant growth and SOC accumulation [41]. The acidity can enhance SOC by slowing organic matter decomposition. EC, which indicates soil salinity and nutrient cycling, typically rises with soil solution content but can impede plant growth and SOC accumulation. The relationship between EC, SOC, and altitude varies, reflecting the ecological complexity of mountain ecosystems [42].

Soil BD showed minimal variation across conditions, with a slight decrease in disturbed areas, as it is influenced by soil structure, texture [43], and organic matter, which decreases with altitude and increases with soil depth. However, climate change and land management practices can exacerbate soil compaction and impact SOC and BD interactions [44]. Lehmann and Kleber [45] found patterns similar to those revealed by this study. Furthermore, the relationship between BD and SOC is explained by Six et al. [46], who demonstrated that compacted soils retain SOC by slowing down microbial decomposition.

Our findings indicated that while disturbances significantly reduced SOM and SOC [8, 29, 47], other soil physical and chemical parameters exhibited resilience to disturbance in the studied mountain ecosystem. This resilience may be vital for maintaining ecosystem functions despite changes in soil organic matter and carbon stocks [10].

Model Evaluation and Applicability

Highly correlated predictors can obscure the influence of individual variables in statistical models. The Variance Inflation Factor (VIF) helps identify these issues, with values above 5 or 10 indicating potential multicollinearity problems [48]. Addressing multicollinearity may involve removing or combining predictors or using regularization techniques. Standardizing variables is crucial for accurate comparisons across different scales, particularly in methods like PCA and BLM, to prevent any single variable from disproportionately influencing the results [49].

RMSE is a key metric to evaluate the accuracy of predictive models, with values ranging from zero to infinity. Lower RMSE values indicate better model performance, which is affected by the model's complexity, the data's distribution, and the presence of outliers [50]. The models were assessed based on their RMSE values. The GAM exhibited the highest RMSE at 20.91, indicating the lowest predictive accuracy among the models evaluated. Following GAM, the BLM recorded an RMSE of 19.1, while the LMM demonstrated an RMSE of 16.52, reflecting a moderate level of predictive capability. Notably, the MRA model achieved an RMSE of 16.26 for SOC predictions and a remarkably low RMSE of 0.61 for pH predictions, signifying exceptional accuracy in pH estimations. The RFR model emerged as the most accurate, with an RMSE of 13.46 for SOC, showing its effectiveness in capturing complex interactions and non-linearities within the dataset.

In this study, flexible models like RFR and MRA outperformed linear models, as evidenced by their lower RMSE values. RFR performed best due to its ability to capture complex, non-linear relationships in the data, while MRA excelled in pH predictions. In contrast, linear models such as GAM, BLM, and LMM had higher RMSE values, with GAM showing the lowest predictive accuracy, likely because linear models resisted more complicated data patterns. The results revealed the advantage of flexible models in handling complex interactions, where linear models fall short [51]. The variations in RMSE among the models also illustrated the influence of model complexity and data structure on predictive accuracy [52].

PCA reduced correlated data to a smaller set of uncorrelated components, explaining 78.35% of the variance with the top three components. The first reflected the role of clay in stabilizing SOC by protecting it from decomposition [53]. Sandy soils, which showed lower loadings, have lower SOC due to reduced protection of soil organic matter. This analysis underscores the significance of soil texture in SOC stabilization [43].

LMMs indicated that soil type significantly influenced SOC ($p<0.001$), while altitude had no notable effect ($p = 0.703$). Similarly, GAM showed no significant non-linear effect of altitude on SOC, highlighting the need to account for other factors like climate and soil composition. This indicates that altitude alone does not fully explain SOC variability, with factors like organic input and erosion playing key roles. BLM showed a slight positive effect of altitude on SOC, with credible intervals pointing to reduced decomposition at higher elevations. Soil texture, particularly loamy sand and sandy loam, influenced SOC retention [43, 54]. RFR emphasized the role of BD in limiting microbial activity and enhancing SOC retention [46]. EC and pH also significantly affected nutrient availability and microbial activity, which is crucial to SOC dynamics [53].

MRA effectively modeled SOC and pH together, demonstrating how factors like altitude and soil type interact to influence both variables. The residuals showed a good fit, affirming the relevance of the predictors. The consistency of results across models supports the robustness of findings, particularly regarding the impact of soil type on SOC, reducing concerns about model-specific biases [31]. The alignment across models ensures reliable predictions and a comprehensive understanding of SOC and pH dynamics [55].

The models linked SOC dynamics to ecological processes, emphasizing the importance of soil type and texture in SOC stabilization. Clay-rich soils were found to better protect SOC from decomposition than sandy soils. While altitude did not directly influence SOC, it likely affects it through associated variables like temperature and moisture [56]. However, the RFR model revealed an improved impact of BD compared to simpler models [57]. However, discrepancies with other studies may arise due to data types used for modeling and local conditions.

Modeling techniques face limitations that can influence outcomes. Missing data, if not random, can introduce bias, and both imputation and data exclusion may affect representativeness and prediction accuracy [58]. Small sample sizes, especially in complex models, can lead to overfitting or underfitting, reducing parameter reliability and model performance. Additionally, limited spatial data or uneven coverage can hinder the ability to accurately capture variability, leading to the misrepresentation of predictor relationships in diverse landscapes [58, 59].

Future soil carbon modeling research could be improved by adopting more advanced models, such as deep learning, to capture non-linear interactions that traditional methods might miss. High-resolution data from remote sensing or frequent soil measurements could increase predictive accuracy, offering a deeper understanding of SOC dynamics [60]. Enhanced cross-validation methods, including factors like microbial activity or land management practices, could further improve model reliability. Validating models with independent datasets across diverse regions would strengthen the applicability of the findings [61].

The practical implications of this study are significant, as it provides information that can influence environmental policy and promote sustainable land management practices. A better understanding of SOC dynamics can guide carbon management strategies and support the development of effective SOC management practices [62]. This is vital for climate models that rely on accurate SOC data to predict carbon fluxes and assess mitigation strategies. The findings can also support conservation efforts by emphasizing the importance of soil texture and bulk density in carbon storage [63]. Together, these improvements could lead to more effective soil and climate management policies.

Conclusions

This study provides valuable insights into SOC dynamics and associated physicochemical properties in disturbed and undisturbed soils across different altitudes. The results highlight the following key findings: (i) Undisturbed soils have significantly higher SOC levels, highlighting the negative impact of disturbances and land degradation. (ii) Altitude had no significant effect on SOC levels in the study area, as supported by models like the LMM and GAM. (iii) Soil properties (pH, EC, and BD) showed resilience to disturbance. BD emerged as a key predictor of SOC retention, particularly in the RFR model, which captured non-linear relationships better than linear models. (iv) PCA reinforced the role of clay-rich soils in SOC stabilization, while sandy soils were linked to lower SOC levels. (v) The RFR model performed best in capturing SOC variability; however, cross-validation is necessary to avoid overfitting. (vi) While consistent with existing research, discrepancies regarding the influence of altitude may stem from regional differences in soil composition and conditions. The study suggests that future research could benefit from incorporating advanced modeling techniques and additional variables, such as microbial activity, to further refine SOC predictions. These findings have practical implications for soil carbon management and climate mitigation, emphasizing the importance of soil structural attributes in carbon sequestration and sustainable land management practices.

Acknowledgments

The authors wish to express their gratitude to the research scholars of the Plant Ecology and Environmental Science Laboratory in the Department of Botany at the University of Azad Jammu and Kashmir for their valuable assistance and support during the fieldwork.

Conflict of Interest

The authors declare no conflict of interest.

References

1. RODRIGUES C.I.D., BRITO L.M., NUNES L.J. Soil carbon sequestration in the context of climate change mitigation: a review. Soil Systems. **7** (3), 64, **2023**.
2. XU M.J., LI G., NAZIR M.M., ZULFIQAR F., SIDDIQUE K.H., IQBAL B., DU D. Harnessing soil carbon sequestration to address climate change challenges in agriculture. Soil and Tillage Research. **237**, 105959, **2024**.
3. DON A., SEIDEL F., LEIFELD J., KÄTTERER T., MARTIN M., PELLERIN S., EMDE D., SEITZ D., CHENU C. Carbon sequestration in soils and climate change mitigation—definitions and pitfalls. Global Change Biology. **30** (1), e16983, **2024**.
4. NEGI V.S., PATHAK R., DHYANI V., DURGAPAL M., JOSHI R., BHATT I.D. Land restoration in the Himalayan Region: Steps towards biosphere integrity. Land Use Policy. **121**, 106317, **2022**.
5. ANAND A., GARG V.K. Temperature–precipitation trends and response of high-altitude biodiversity reserve of western Himalayas. Journal of Earth System Science. **133** (2), 76, **2024**.
6. AUGUSTO L., BOČA A. Tree functional traits, forest biomass, and tree species diversity interact with site properties to drive forest soil carbon. *Nature Communications*. **13** (1), 1, **2022**.
7. KHANAL S., NOLAN R.H., MEDLYN B.E., BOER M.M. Environmental correlates of the forest carbon distribution in the Central Himalayas. *Ecology and Evolution*. **14** (6), e11517, **2024**.
8. QIU L., ZHU H., LIU J., YAO Y., WANG X., RONG G., ZHAO X., SHAO M., WEI X. Soil erosion significantly reduces organic carbon and nitrogen mineralization in a simulated experiment. Agriculture, Ecosystems & Environment. **307**, 107232, **2021**.
9. YANG Y., QIU K., XIE Y., LI X., ZHANG S., LIU W., HUANG Y., CUI L., WANG S., BAO P. Geographical, climatic, and soil factors control the altitudinal pattern of rhizosphere microbial diversity and its driving effect on root zone soil multifunctionality in mountain ecosystems. Science of The Total Environment. **904**, 166932, **2023**.
10. GAO H., GONG J., LIU J., YE T. Effects of land use/cover changes on soil organic carbon stocks in Qinghai-Tibet plateau: A comparative analysis of different ecological functional areas based on machine learning methods and soil carbon pool data. Journal of Cleaner Production. **434**, 139854, **2024**.
11. THIAGARAJAN A., LIANG C., MACDONALD J.D., SMITH W., VANDENBYGAART A., GRANT B., KROBEL R., JANZEN H., ZHANG T., MCCONKEY B., MA B., BREMER E., YANG X., CERKOWNIAK D., FAN J. Prospects and challenges in the use of models to estimate the influence of crop residue input on soil organic carbon in long-term experiments in Canada. Geoderma Regional. **30**, e00534, **2022**.
12. LI T., CUI L., WU Y., MCLAREN T.I., XIA A., PANDEY R., LIU H., WANG W., XU Z., SONG X., DALAL R.C., DANG Y.P. Soil organic carbon estimation via remote sensing and machine learning techniques: Global topic modeling and research trend exploration. Remote Sensing. **16** (17), 3168, **2023**.
13. YUZUGULLU O., FAJRAOUI N., DON A., LIEBISCH F. Satellite-based soil organic carbon mapping on European soils using available datasets and support sampling. Science of Remote Sensing. **9**, 100118, **2024**.
14. BASHIR O., BANGROO S.A., SHAFAI S.S., SENESI N., KADER S., ALAMRI S. Geostatistical modeling approach for studying total soil nitrogen and phosphorus under various land uses of North-Western Himalayas. Ecological Informatics. **80**, 102520, **2024**.
15. KHAN A.M., QURESHI R., SAQIB Z. Multivariate analyses of the vegetation of the western Himalayan forests of Muzaffarabad district, Azad Jammu and Kashmir, Pakistan. Ecological Indicators. **104**, 723, **2019**.
16. WWF-PAKISTAN. Annual Report. Available online: https://d2ouvy59p0dg6k.cloudfront.net/downloads/annual_report_2019_web.pdf (Accessed on 12 08 **2024**).
17. UMAR M., HUSSAIN M., MURTAZA G., SHAHEEN F.A., ZAFAR F. Ecological concerns of migratory birds in Pakistan: a review. Punjab University Journal of Zoology. **33** (1), 69, **2018**.
18. WANI Z.A., NEGI V.S., BHAT J.A., SATISH K.V., KUMAR A., KHAN S., DHYANI R., SIDDIQUI S., Al-QTHANIN R.N., PANT S. Elevation, aspect, and habitat heterogeneity determine plant diversity and compositional patterns in the Kashmir Himalaya. Frontiers in Forests and Global Change. **6**, 1019277, **2023**.
19. VINODHA K., GOPI E.S., BHAGWANSWAROOP B.D., KUMAR A.P. Generalized soil texture classification using systematic generation of mixtures based on USDA soil classification triangle. IEEE Geoscience and Remote Sensing Letters. **2024**.
20. BLACK G.R., HARTGE K.H. Bulk density 1. Method of soil analysis: part 1- Soil Science Society of American Journal. **5**, 363, **1986**.
21. WALKLEY A., BLACK I.A. An examination of Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science. **37** (1), 29, **1934**.
22. SÁ J DE M., CERRI C.C., DICK W.A., LAL R., VENSKE FILHO S.P., PICCOLO M.C., FEIGL B.E. Organic matter dynamics and carbon sequestration rates for a tillage chronosequence in a Brazilian Oxisol. Soil Science Society of America Journal. **65** (5), 1486, **2001**.
23. SONG J., YU D., WANG S., ZHAO Y., WANG X., MA L., LI J. Mapping soil organic matter in cultivated land based on multi-year composite images on monthly time scales. Journal of Integrative Agriculture. **23** (4), 1393, **2024**.
24. ROBACK P., LEGLER J. Beyond multiple linear regression: Applied generalized linear models and multilevel models in R. Chapman and Hall/CRC, **2021**.
25. R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, **2024**.
26. CAI H., WANG C., MA Z., MENG F., LIN Z., REN J., LI S. Predicting frost heave in soil-water systems using the generalized regression neural network optimized with particle swarm optimization algorithm. Cold Regions Science and Technology. **226**, 104291, **2024**.
27. HAN H., SUH J. Spatial prediction of soil contaminants using a hybrid random forest–ordinary kriging model. Applied Sciences. **14** (4), 1666, **2024**.
28. ZHOU W., LI C., ZHAO W., STRINGER L.C., FU B. Spatial distributions of soil nutrients affected by land use, topography and their interactions, in the Loess Plateau of China. International Soil and Water Conservation Research. **12** (1), 227, **2024**.
29. SAINEPO B.M., GACHENE C.K., KARUMA A. Assessment of soil organic carbon fractions and carbon management index under different land use types in

Olesharo Catchment, Narok County, Kenya. Carbon Balance and Management. **13**, 1, **2018**.

30. ZHAO Z., GENG P., WANG X., LI X., CAI P., ZHAN X., HAN X. Improvement of active organic carbon distribution and soil quality with the combination of deep tillage and no-tillage straw returning mode. Agronomy. **13** (9), 2398, **2023**.
31. LAL R. Soil carbon sequestration to mitigate climate change. Geoderma. **123** (1-2), 1, **2004**.
32. SHEIKH M.A., KUMAR M., BUSSMANN R.W. Altitudinal variation in soil organic carbon stock in coniferous subtropical and broadleaf temperate forests in Garhwal Himalaya. Carbon Balance and Management. **4**, 1, **2009**.
33. PANWAR V.P., GUPTA M. Soil organic carbon pool under different forest types in Himachal Pradesh. International Journal of Farm Sciences. **3** (2), 81, **2013**.
34. SING P.P., RAWAT Y.S. Altitude wise variation in soil carbon stock in Western Himalaya. New York Scientific Journal. **6** (8), 43, **2013**.
35. ARORA G., CHATURVEDI S., KAUSHAL R., NAIN A., TEWARI S., ALAM N.M., CHATURVEDI O.P. Growth, biomass, carbon stocks, and sequestration in an age series of Populus deltoides plantations in Tarai region of central Himalaya. Turkish Journal of Agriculture and Forestry. **38**, 550, **2014**.
36. SAEED S., ASHRAF M.I., AHMAD A., RAHMAN Z. The Bela forest ecosystem of district Jhelum, a potential carbon sink. Pakistan Journal of Botany. **48** (1), 121, **2016**.
37. SHAHEEN H., KHAN R.W.A., HUSSAIN K., ULLAH T.S., NASIR M., MEHMOOD A. Carbon stocks assessment in subtropical forest types of Kashmir Himalayas. Pakistan Journal of Botany. **48** (6), 2351, **2016**.
38. AZIZ S., CHUGHTAI F.M., SHAHEEN H., KHAN R.W.A., DAR M.E.U.I. Biomass and soil carbon stocks assessment in western Himalayan alpine and subalpine vegetation zones of Kashmir. Pakistan Journal of Botany. **51**, 973, **2019**.
39. SIMON A., DHENDUP K., RAI P., GRATZER G. Soil carbon stocks along elevational gradients in Eastern Himalayan mountain forests. Geoderma Regional. **12**, 28, **2018**.
40. YANG D., HUANG X., SHE D., FANG N., NI L., SHI Z. Revegetation of sloping land significantly reduces SOC loss via erosion on the Loess Plateau. Agriculture, Ecosystems & Environment. **376**, 109225, **2024**.
41. MASSACCESI L., DE FEUDIS M., LECCESE A., AGNELLI A. Altitude and vegetation affect soil organic carbon, basal respiration and microbial biomass in Apennine forest soils. Forests. **11** (6), 710, **2020**.
42. HASSANI A., SMITH P., SHOKRI N. Negative correlation between soil salinity and soil organic carbon variability. Proceedings of the National Academy of Sciences of the United States of America. **121** (18), e2317332121, **2024**.
43. PANAGOS P., ROSA D.D., LIAKOS L., LABOUYRIE M., BORRELLI P., BALLABIO C. Soil bulk density assessment in Europe. Agriculture, Ecosystems & Environment. **364**, 108907, **2024**.
44. HAMIDOV A., HELMING K., BELLOCCHI G., BOJAR W., DALGAARD T., GHALEY B.B., HOFFMANN C., HOLMAN I., HOLZKÄMPER A., KRZEMINSKA D., KVÆRNØ S.H., LEHTONEN H., NIEDRIST G., ØYGARDEN L., REIDSMA P., ROGGERO P.P., RUSU T., SANTOS C., SEDDAIU G., SKARBØVIK E., VENTRELLA D., ŻARSKI J., SCHÖNHART M. Impacts of climate change adaptation options on soil functions: a review of European case-studies. Land Degradation & Development. **29** (8), 2378, **2018**.
45. LEHMANN J., KLEBER M. The contentious nature of soil organic matter. Nature. **528** (7580), 60, **2015**.
46. SIX J., CONANT R.T., PAUL E.A., PAUSTIAN K. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant and Soil. **241** (2), 155, **2002**.
47. CURRAN M.P., MURRAY M.P. Soil disturbance, amelioration and rehabilitation affect forest growth, health, soil carbon and chemistry on five long-term soil productivity (LTSP) sites in southeastern British Columbia. Forest Ecology and Management. **546**, 121362, **2023**.
48. DERTLI H.I., HAYES D.B., ZORN T.G. Effects of multicollinearity and data granularity on regression models of stream temperature. Journal of Hydrology. **639**, 131572, **2024**.
49. SHRESTHA N. Detecting multicollinearity in regression analysis. American Journal of Applied Mathematics and Statistics. **8** (2), 39, **2020**.
50. KOUZEHGAR K., ESLAMIAN S. Application of experimental data and soft computing techniques in determining the outflow and breach characteristics in embankments and landslide dams. In Handbook of Hydroinformatics, pp.11-31, **2022**.
51. MCCABE C.J., HALVORSON M.A., KING K.M., CAO X., KIM D.S. Interpreting interaction effects in generalized linear models of nonlinear probabilities and counts. Multivariate Behavioral Research. **57** (2-3), 243, **2021**.
52. ZIOLKOWSKI P. Computational complexity and its influence on predictive capabilities of machine learning models for concrete mix design. Materials. **16** (17), 5956, **2023**.
53. DAS A., PURAKAYASTHA T.J., AHMED N., DAS R., BISWAS S., SHIVAY Y.S., SEHGAL V.K., RANI K., TRIVEDI A., TIGGA P., SAHOO J., CHAKRABORTY R., SEN S. Influence of clay mineralogy on soil organic carbon stabilization under tropical climate, India. Journal of Soil Science and Plant Nutrition. **23** (1), 1003, **2023**.
54. WIESMEIER M., URBANSKI L., HOBLEY E., LANG B., VON LÜTZOW M., MARIN-SPIOTTA E., VAN WESEMAEL B., RABOT E., LIEß M., GARCIA-FRANCO N., WOLLSCHLÄGER U., VOGEL H., KÖGEL-KNABNER I. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. Geoderma. **333**, 149, **2018**.
55. ADENIYI O.D., BRENNING A., MAERKER M. Spatial prediction of soil organic carbon: combining machine learning with residual kriging in an agricultural lowland area (Lombardy region, Italy). Geoderma., **448**, 116953, **2024**.
56. SEKAR K.C., THAPLIYAL N., PANDEY A., JOSHI B., MUKHERJEE S., BHOJAK P., BAHUKHANDI A. Plant species diversity and density patterns along altitude gradient covering high-altitude alpine regions of west Himalaya, India. Geology, Ecology, and Landscapes. **8** (3), 1, **2023**.
57. LUAN J., ZHANG C., XU B., XUE Y., REN Y. The predictive performances of random forest models with limited sample size and different species traits. Fisheries Research. **227**, 105534, **2020**.
58. ZINN S., GNAMBS T. Modeling competence development in the presence of selection bias. Behavior

Research Methods. **50** (6), 2426, **2018**.
59. MONTESINOS LÓPEZ O.A., MONTESINOS LÓPEZ A., CROSSA J. Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction, pp. 109. Cham: Springer International Publishing, **2022**.
60. LIU L., ZHOU W., GUAN K., PENG B., XU S., TANG J., ZHU Q., TILL J., JIA X., JIANG C., WANG S., QIN Z., KONG H., GRANT R., MEZBAHUDDIN S., KUMAR V., JIN Z. Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. Nature Communications. **15** (1), 357, **2024**.
61. KALIAPPAN J., BAGEPALLI A.R., ALMAL S., MISHRA R., HU C., SRINIVASAN K. Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise. Diagnostics. **13** (10), 1692, **2023**.
62. CHALCHISSA F.B., KURIS B.K. Modelling soil organic carbon dynamics under extreme climate and land use and land cover changes in Western Oromia Regional State, Ethiopia. Journal of Environmental Management. **350**, 119598, **2024**.
63. TOPA D., CARA I.G., JITĂREANU G. Long term impact of different tillage systems on carbon pools and stocks, soil bulk density, aggregation and nutrients: A field meta-analysis. CATENA. **199**, 105102, **2021**.