INTERNATIONAL
Agrophysics
www.international-agrophysics.org

Int. Agrophys., 2025, 39, 443-456
doi: 10.31545/intagr/206078

Interpretable machine learning unravels hierarchical controls on soil organic carbon in hyperarid oases of the Eastern Tarim Basin**

Jingyu Liu[1]*, *Yuhong Liu*[2], *Junling He*[1], *Shuai Guo*[1], *Pengwei Zhang*[1], *Liqi Meng*[1], *Haoshuai Niu*[1], *Chen Li*[3], *Wentong Xu*[4], *Liming Liu*[5]

[1]Urumqi Natural Resources Comprehensive Survey Center, China Geological Survey, Urumqi, 555 Xihuan North Road, 830057, China
[2]Baotou Teachers College, Baotou, 24 Ziyou Road, 014030, China
[3]Hami Geological Brigade, Geological Bureau of Xinjiang Uygur Autonomous Region, Hami, 38 Jianguo North Road, 839099, China
[4]Bayingolin Geological Brigade, Geological Bureau of Xinjiang Uygur Autonomous Region, Korla, 83 West Tianshan Road, 841000, China
[5]Hubei Provincial Land and Resources Information Center, Geological Bureau of Hubei Province, Wuhan, 27, Gongzheng Road, 430071, China

Received August 12, 2024; accepted January 21, 2025

Abstract. Understanding spatial drivers of soil organic carbon in arid oasis ecosystems is essential for guiding precision soil management and enhancing land sustainability. This study integrates 644 surface samples and 9 soil profiles with multi-source environmental data in the hyperarid eastern Tarim Basin, employing geostatistics and machine learning (random forest, support vector machines, ordinary least squares, back propagation) to quantify driving mechanisms. Key findings: 1) extreme soil organic carbon spatial polarization (0.50-21.70 g kg^{-1}, mean = 4.47 g kg^{-1}) with northern and southern alluvial zones containing 2.1 times higher soil organic carbon than central deserts ($p < 0.01$); 2) random Forest achieved optimal prediction (coefficient of determination = 0.81, root mean square error = 1.32 g kg^{-1}) by resolving nonlinear soil organic carbon-environment interactions; 3) pedogenic properties (texture, cation exchange capacity, salinity; 47.5%) dominated soil organic carbon variation, followed by anthropogenic drivers (land use intensity, 14.2%) and soil taxonomy (10.9%), while climate and topography showed minimal control (< 8%). Human-modified processes override climatic constraints in shaping soil organic carbon patterns, providing actionable insights for clay-organic stabilization and irrigation optimization. This methodology establishes a transferable framework for deciphering soil organic carbon dynamics in global drylands, directly informing climate-resilient land management.

Keywords: carbon sequestration, dryland soils, geospatial analysis, predictive modeling, sustainable agriculture

*Corresponding author e-mail: 965690970@qq.com
**This work was founded by the Key Laboratory of Coupling Process and Effect of Natural Resources Elements (No.2024KFKT019) Geological Survey Projects of the China Geological Survey (DD20242035) (2024-2026).

1. INTRODUCTION

As a pivotal carbon reservoir in terrestrial ecosystems, the pedosphere (soil sphere) stores approximately three-quarters of global organic carbon, playing a critical role in Earth's carbon cycle (Neupane *et al*., 2022). Notably, soil organic carbon (SOC) storage exceeds the combined carbon stocks in atmospheric and phytomass pools. The dynamic equilibrium of SOC not only exhibits strong coupling with global climate patterns and atmospheric CO_2 concentrations but also serves as a key regulator of soil health and biodiversity conservation (Wang Z. *et al*., 2024; Annalisa *et al*., 2024). The spatial heterogeneity of SOC content arises from the complex interplay between carbon input-output budgets and environmental drivers, resulting in non-stationary distribution patterns with significant scale dependence (Wang L. *et al*., 2024). Understanding these spatial dynamics and their governing mechanisms carries profound implications for soil quality enhancement, agricultural optimization, and climate change mitigation (Boubehziz *et al*., 2024).

SOC dynamics are governed by two fundamental dimensions: stochastic environmental forcing and intrinsic soil properties. External stochastic disturbances include such factors as human activities (farming, burning) and environmental changes (rainfall, temperature). These factors

influence SOC content by altering the redistribution of water and heat resources during soil formation processes, as well as through erosion effects (Qin *et al*., 2024; Xiong *et al*., 2024; Szostek *et al*., 2022; Noppol *et al*., 2021). The latter involves edaphic characteristics including soil texture, parent material composition, and physicochemical properties that regulate microbial activity and soil structural stability (Xia *et al*., 2024; Qin *et al*., 2020). Empirical evidence suggests that parent material exerts dominant control on SOC spatial heterogeneity in agricultural soils of southern China's hilly regions (Zhang *et al*., 2021), while macroscale patterns are predominantly shaped by climate-topography interactions through soil moisture modulation (Li S. *et al*., 2024; Li X.L. *et al*., 2024). Furthermore, SOC demonstrates textural dependency, with positive correlations with clay/silt fractions and negative associations with sand content (Sanleandro *et al*., 2023). Despite advances in quantifying SOC drivers and spatial variability, critical knowledge gaps persist regarding the hierarchical controls and causal pathways underlying SOC distribution, particularly across spatial scales (Huyzentruyt *et al*., 2024; Spotorno *et al*., 2024). This limitation constrains effective land management strategies, as environmental stochasticity (thermal-hydrological regimes) gains prominence at broader scales (Yang *et al*., 2021).

The Tarim Basin, a global archetype of arid ecosystems, presents unique pedological challenges characterized by pronounced soil degradation, suboptimal land use efficiency, and SOC spatial disequilibrium (Cao *et al*., 2024). Its distinctive geoclimatic setting marked by extreme aridity, complex topography, and oasis-desert ecotones provides an ideal natural laboratory for investigating SOC dynamics in water-limited conditions. Focusing on the eastern marginal oasis of this basin, our study integrates soil parameter analysis with geostatistical modeling and machine learning approaches to: 1) delineate SOC spatial distribution patterns, and 2) quantify the relative contributions of environmental and edaphic drivers. This multi-method framework advances the understanding of SOC sequestration mechanisms in arid ecosystems while informing sustainable land management strategies for analogous regions (Yin *et al*., 2012). Our findings elucidate the spatial organization of SOC stocks and establish a causal hierarchy among controlling factors, providing actionable insights for soil ecological restoration and climate-smart agriculture in dryland environments.

2. MATERIALS AND METHODS

2.1. Overview of the study area

The study area (86°59′-88°02′E, 40°55′-41°09′N) spans the ecotone between the southern Tianshan piedmont and eastern Tarim Basin in northwestern China (Fig. 1). Encompassing 2,215.52 km^2 with maximum east-west and north-south extents of 121.65 km and 61.78 km respectively, this transitional zone exhibits an elevational gradient from 800 to 1 200 m asl. The terrain displays subdued relief with a gentle southwesterly slope, manifesting higher elevations in the northern and eastern sectors relative to the southern and western lowlands.

Climatically, the region is classified as a warm temperate continental arid zone, with pronounced orographic effects from the Tianshan Mountains and Gobi Desert creating extreme aridity. The mean annual temperature reaches 11.4°C, while the hydrologic regime is characterized by a severe moisture deficit – annual potential evaporation exceeds precipitation by two orders of magnitude (≈100:1 ratio). The ephemeral Tarim River constitutes the principal drainage feature, traversing the study area along

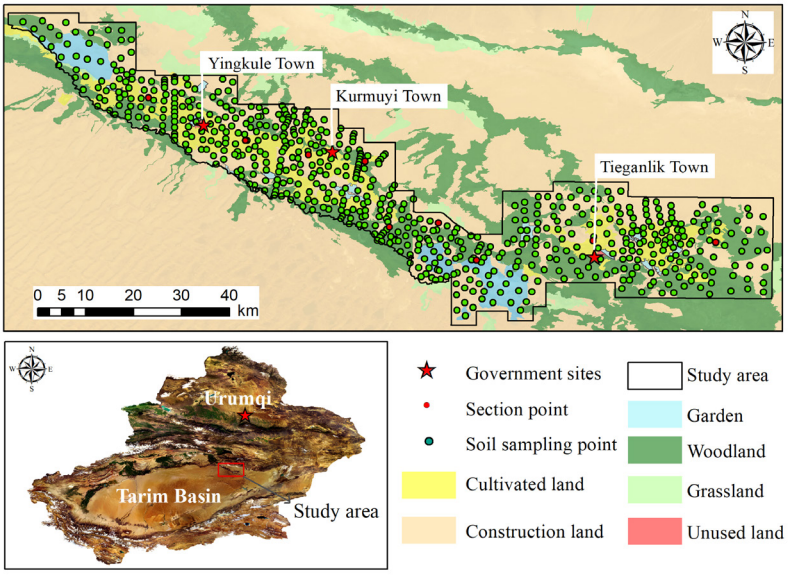

Fig. 1. Geographical location of the study area and spatial distribution of sampling points.

a NW-SE axis. Pedogenic processes in this desert-oasis ecotone yield four dominant soil types: oasis loess, brown calcareous soils, saline soils, and aeolian sands. Xeric vegetation assemblages feature halophytic shrubs including *Tamarix chinensis*, *Populus euphratica*, and *Alhagi sparsifolia* exhibiting adaptations to hyperarid conditions. The anthropogenic land use centers on intensively irrigated agriculture, with cash crops (cotton, jujube, walnut, pear) constituting primary cultivation systems.

Recent development of the Tarim Economic Belt has accentuated the region's agroeconomic potential, driving the demand for enhanced soil physicochemical optimization in precision farming systems (Yin *et al*., 2012). However, agricultural intensification has precipitated significant pedospheric perturbations. Documented impacts include acute land use conflicts, substantive soil quality decline (Liu *et al*., 2016), and SOC depletion from non-sustainable practices like excessive tillage and fertilizer misuse (Huong *et al*., 2022). These anthropogenic stressors collectively threaten both pedological health and long-term agricultural viability in this fragile arid ecosystem.

2.2. Sample collection and handling

A stratified sampling strategy was implemented using a grid-based design (1/4 km^2 resolution), accounting for spatial heterogeneity in land use patterns, soil textural classes, and salinity gradients. This systematic approach yielded 644 georeferenced surface soil sampling points, with field adjustments implemented to accommodate logistical constraints. To characterize the vertical SOC distribution, nine representative pedons were strategically established across the study area, each sampled at seven depth increments (0-210 cm) with 30 cm intervals. Field protocols mandated removal of extraneous biogenic materials (roots, gravel > 2 mm) followed by homogenization of composite samples. Quadruplicate aliquots (~1 kg each) were sealed in sterile containers and transported in ventilated conditions. Laboratory processing involved air-drying, disaggregation, and sieving (< 850 μm, 20-mesh), with concurrent documentation of pedological characteristics (soil taxonomy, land management history) and environmental covariates.

A total of eight indicators were analyzed and tested, including SOC, total potassium (TK), total nitrogen (TN), total phosphorus (TP), total salinity (TS), pH, bulk density (BD), and soil texture. The testing methods, instrumentation, and detection limits for each indicator are presented in Table 1. Rigorous quality assurance protocols were implemented, including procedural blanks, analytical duplicates (10% of samples), and certified reference materials (NIST 2709a). Daily instrument calibration utilized matrix-matched standards, with all measurements conforming to Method Quantification and Control Limits (MQCL) and demonstrating < 5% relative standard deviation in replicate analyses.

2.3. Research methodology

2.3.1. Statistical analysis method

The study integrated continuous (topography, climate, soil physicochemical properties) and categorical variables (soil texture types, classification categories) to assess the drivers of SOC dynamics. Pearson/Spearman correlation analysis quantified relationships between SOC and continuous variables, while ANOVA evaluated spatial heterogeneity across categorical groups. A multivariate linear regression framework modeled interactions among variables influencing oasis SOC at the Tarim Basin's eastern margin, incorporating stepwise selection to isolate causal factors. Qualitative variables were dummy-coded prior to regression, with model diagnostics including Durbin-Watson autocorrelation tests (residual independence), R^2-adjusted variance partitioning, and partial/semi-partial correlation coefficients to disentangle covariate effects. Covariance structures were validated through variance

Table 1. Analytical testing methods and instrument equipment

Element	Method	Instrument	Detection limit
SOC	Walkley-Black method	Evolution One spectrophotometer (USA)	0.3 g kg^{-1}
TK	Inductively coupled plasma optical emission spectrometry	NexION 300D inductively coupled plasma mass spectrometer (USA)	0.1 g kg^{-1}
TN	Kjeldahl digestion-steam distillation method	Kjeltec™ 9 fully automated Kjeldahl nitrogen analyzer (Denmark)	0.1 g kg^{-1}
TP	X-ray fluorescence spectrometry	AXIOS MAX X-ray fluorescence spectrometer (Netherlands)	0.05 g kg^{-1}
TS	Gravimetric method	BSA124S-CW electronic balance (Germany)	0.5 g kg^{-1}
pH	Ion-selective electrode potentiometry	S220 ion meter (Switzerland)	0.01
BD	Core sampling method	100 cm^3 standard cutting ring (China)	0.01 g cm^{-3}
Soil texture	Hydrometer method	Standard hydrometer (China)	0.10%

inflation factors (VIF < 5). All computations were executed in SPSS 26 (α = 0.05), with spatial trend visualizations generated in Origin 2022.

2.3.2. Spatial geostatistical methods

Spatial heterogeneity of SOC in the Tarim Basin's eastern margin oasis was quantified using geostatistical modeling in GS+ software. Four theoretical semivariogram models (Gaussian, linear, exponential, spherical) were iteratively fitted to empirical semivariance data, with model performance evaluated through dual criteria: maximization of the coefficient of determination (R^2) and minimization of the residual sum of squares (RSS). The optimal model, selected based on these metrics, was subsequently applied in ArcGIS 10.2 to execute ordinary kriging interpolation, generating spatially continuous SOC predictions across the study area. Prediction uncertainty was assessed via cross-validation, calculating mean error (ME) between observed SOC values and model estimates at sampling locations. This framework explicitly accounts for spatial autocorrelation while mapping SOC distribution patterns at landscape scales.

2.3.3. Machine learning methods

1) Ordinary least squares (OLS)

As a fundamental linear modeling technique, OLS regression was chosen for its interpretability, computational efficiency, and well-established theoretical framework (Mackessy *et al*., 2024). While nonlinear models offer flexibility, OLS serves as a critical baseline for evaluating linear relationships, particularly when preliminary analyses or domain knowledge suggest approximate linearity in the data. Its simplicity facilitates rapid diagnostics and benchmarking against more complex alternatives. The mathematical formulation of the OLS regression model is as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n + \varepsilon, \tag{1}$$

where: y is the dependent variable, x_1, x_2, …, x_n denote the independent variables, b_1, b_2, …, b_n. b_n represents the parameters of the model and ε represents the error term.

2) Random forest (RF)

RF is an advancement of the traditional classification and regression tree (CART) methodology, noted for its high predictive efficiency and robustness against overfitting (Liu *et al*., 2024a). By utilizing the bootstrap resampling technique, RF generates multiple subsets of equal size from the original training dataset, with each subset used to train an individual decision tree. For each decision tree, the optimal variable within the subsample is identified for node segmentation. The final prediction output of the RF model is obtained by averaging the predictions from all individual trees, as represented by the following formula:

$$F(x) = \frac{1}{n}\sum_{t=1}^{n} h_t(x), \tag{2}$$

where: $F(x)$ is the final prediction result of RF and $h_t(x)$ is the regression prediction result of the *t-th* decision tree. In this study, the number of decision trees was set to 500, and the minimum leaf subtree was set to 5.

3) Support vector machines (SVM)

The SVM methodology is particularly effective in addressing practical challenges associated with classification and regression tasks, including issues related to small sample sizes, nonlinearity, high-dimensional data, and local minima (Zhang *et al*., 2022). It significantly mitigates the challenges posed by the "curse of dimensionality". The kernel function facilitates nonlinear mapping, transforming the original space into a high-dimensional space, thereby enabling linear classification and regression of samples within this elevated dimensionality. In the context of regression, the resulting model is known as a SVM regression model. The SVM approach constructs an optimal classification hyperplane within the feature space and employs a limited number of support vectors to represent the entire sample space. This strategy enhances the simplicity and robustness of the algorithm when determining the final decision function. The functional representation of this method is as follows:

$$f(x) = \sum_{i=1}^{n}(\hat{a}_i - a_i)x_i^T x_j + b. \tag{3}$$

Introducing the kernel function and substituting $\Phi(x)$ for x, the final model with the kernel function is obtained as follows:

$$f(x) = \sum_{i=1}^{n}(\hat{a}_i - a_i)K(x_i^T x) + b, \tag{4}$$

where: $K(x_i^T x)=\Phi(x_i)^T\Phi(x_j)$, a_i represents the corresponding Lagrange multiplier, n represents the number of samples in the training set, and b represents the offset. The specific parameter is set as penalty factor 50, the radial basis function parameter is 0.2, the loss parameter type is Gaussian SVM, and the parameter is 0.1.

4) Back propagation (BP)

The primary concept of the BP algorithm involves dividing the learning process into two distinct stages: forward propagation and backward propagation (Liu *et al*., 2024). During the forward propagation phase, input samples are processed sequentially from the input layer through the hidden layers, ultimately reaching the output layer after traversing all hidden layers. In this layer-by-layer processing, the state of each layer of neurons influences only the state of the subsequent layer. In the output layer, the current output is compared to the expected output; if a discrepancy exists, the process transitions to backward propagation. In the backpropagation phase, the error between the actual output and the network output is propagated backward along the original connection path.

This error is minimized by adjusting the connection weights of the neurons in each layer. Subsequently, the forward propagation process is repeated, and the calculations continue until the error falls below a predetermined threshold.

The transfer function used in the hidden layer is the hyperbolic tangent sigmoid function (tansig), while the output layer employs the logistic sigmoid function (logsig). The expansion constant for the neural network is set to 1.0. The dynamic adaptive BP algorithm is implemented as the training function, with the stopping criteria defined by target error accuracy and the number of training epochs. The training error accuracy is established at 1×10^{-6} to meet the specified condition, and the maximum number of training iterations is determined by the mathematical formulations for the tansig and logsig functions, as provided below:

$$\text{tansig}(x) = \frac{2}{1+e^{2x}} - 1, \quad (5)$$

$$\text{logsig}(x) = \frac{1}{1+e^{-x}}. \quad (6)$$

2.4. Accuracy evaluation

In this study, three indicators are utilized to evaluate the accuracy of machine learning predictions: the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) (Ma R. *et al*., 2024). R^2 is a widely used metric in statistics that describes the degree of influence of the independent variable on the dependent variable. A value of R^2 closer to 1 indicates that the model fits the sample data more effectively, while a value closer to 0 suggests a poor fit. RMSE and MAE are common metrics for assessing accuracy in machine learning. RMSE is particularly sensitive to large errors, whereas MAE provides a more balanced view of the differences across all error points. Both metrics measure the discrepancy between predicted and actual results, with smaller values indicating better predictive performance of the model (Abdellafou *et al*., 2025; Batista *et al*., 2025; Fu *et al*., 2024).

3. RESULTS

3.1. Distribution characteristics of organic carbon content in the soil surface layer

The surface soil SOC content in the study area ranged from 0.50 to 21.70 g kg^{-1}, with a mean of 4.47 g kg^{-1}, slightly exceeding the 4.36 g kg^{-1} reported for the northern margin of the Tarim Basin but remaining lower than the 7.32 g kg^{-1} observed in Urumqi. The distribution of SOC exhibited pronounced asymmetry (skewness > 1) and leptokurtic peakedness (kurtosis >1), indicating localized enrichment at specific sampling sites. A high coefficient of variation (CV = 65.50%) further underscores strong spatial heterogeneity in SOC distribution (Table 2).

The mean concentrations of TN, TP, and TK were 0.41, 0.28, and 29.3 g·kg^{-1}, respectively. Both TN and TP levels fell below China's urban soil benchmarks (1.07 g kg^{-1} for TN and 0.84 g kg^{-1} for TP), while TK significantly exceeded the national benchmark of 22.80 g kg^{-1}, defining the region's soils as nitrogen-deficient, phosphorus-poor, and potassium-rich (Liu *et al*., 2023).

Soil salinity displayed extreme spatial variability, with a CV of 164.61%, reflecting sharp contrasts in salt accumulation across the study area. In contrast, soil pH exhibited minimal variation, with a mean of 8.65 (strongly alkaline), a low CV of 3.27%, and near-symmetrical distribution (skewness = 0.26, kurtosis = 0.02). This stability in pH contrasts markedly with the high variability observed in SOC and salinity, suggesting distinct geochemical controls on these parameters.

3.2. Characteristics of soil organic carbon vertical distribution

Within the study area, nine soil profiles (PM1-PM9) were established across major land use types: PM1 and PM8 in wasteland, PM2, PM3, and PM9 in arable land, PM4 and PM5 in garden land, PM6 in woodland, and PM7 in grassland. As illustrated in Fig. 2, significant variations in the SOC content were observed among the different land use types. The surface SOC concentrations in arable land (PM2, PM3, PM9) were 10.00, 8.80, and 10.95 g kg^{-1}, respectively, while those in PM1 (wasteland), PM4 and

Table 2. Geochemical characterization of arid oasis soils: statistical summary from the Eastern Tarim Basin Margin (n = 644)

Index	Minimum	Maximum	Average	Standard deviation	Coefficient of variation (%)	Skewness	Kurtosis
TK	12.3	29.3	21.78	2.23	10.26	-0.05	1.82
TN	0.1	1.3	0.41	0.2	48.81	0.97	1.1
SOC	0.5	21.7	4.47	2.84	65.5	1.09	1.62
TP	0.28	1.21	0.58	0.12	20.69	1.37	3.57
TS	0.26	393	23.82	39.21	164.61	3.94	25.41
pH	7.92	9.52	8.65	0.28	3.27	0.26	0.02

TK – total potassium, TN – total nitrogen, SOC – soil organic carbon, TP – total phosphorus, TS – total salinity, pH – potential of hydrogen. Units for TN, TK, SOC, TP, and TS are g kg^{-1}, while pH is dimensionless.

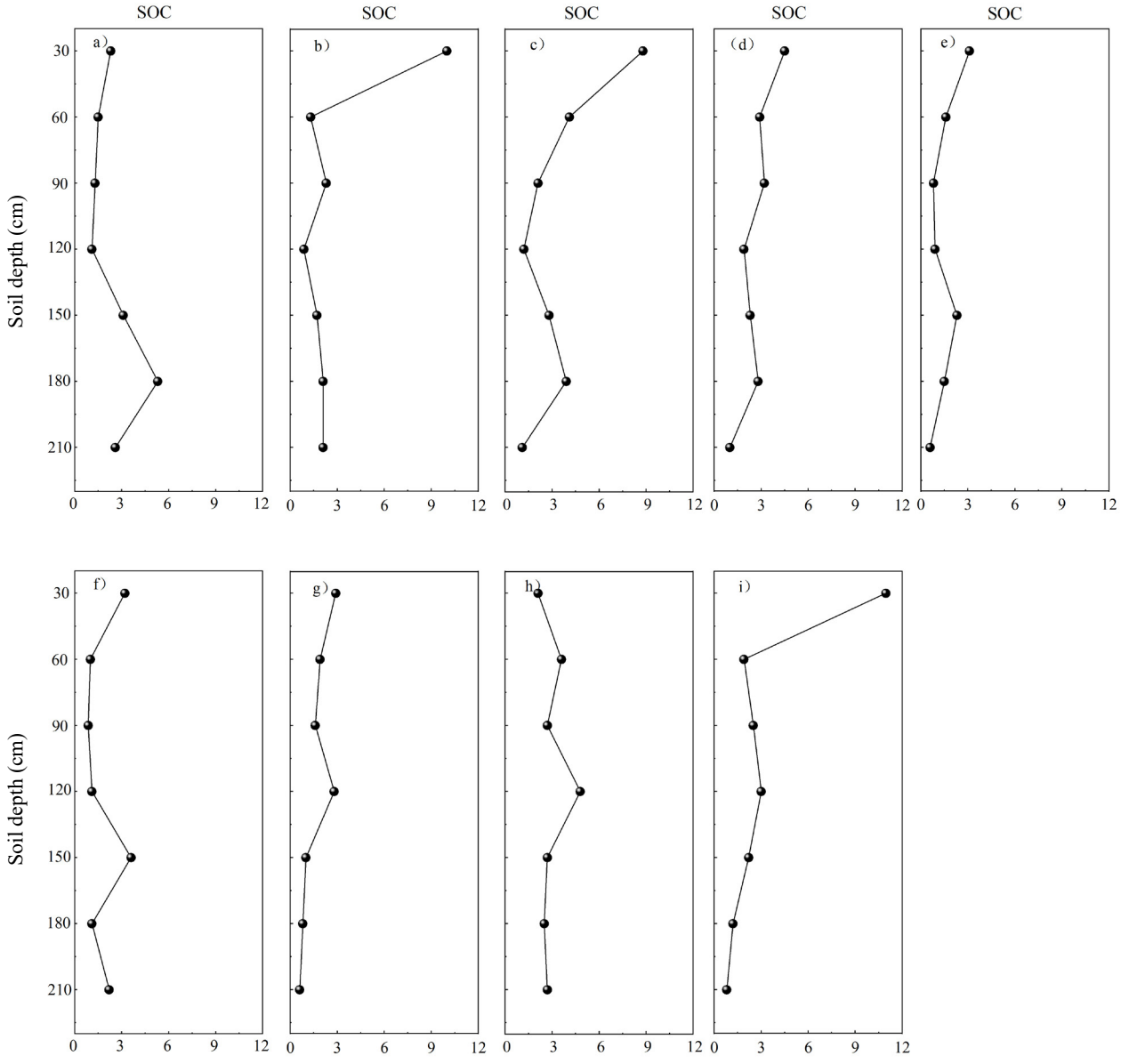

Fig. 2. Vertical stratification of soil organic carbon: depth-specific variability and profile characteristics. Figures a) to i) correspond to PM1 to PM9, respectively.

PM5 (garden land), PM6 (woodland), PM7 (grassland), and PM8 (wasteland) were 2.3, 4.5, 3.5, 3.2, 3.0, and 2.1 g kg^{-1}, respectively. The results indicate that the SOC content exhibited relatively consistent trends within the same land use type but varied significantly across the different types. Notably, the cropland and orchard soils displayed markedly higher SOC levels than the other land use categories, suggesting that anthropogenic agricultural activities (tillage, fertilization, irrigation) significantly influence the migration and accumulation of SOC in the region.

Additionally, except for wasteland (PM4, PM5), the SOC content consistently decreased with increasing soil depth, demonstrating a distinct "surface accumulation" pattern. This phenomenon implies that vegetation root systems facilitate the upward transport of SOC from deeper soil layers during water and nutrient uptake. In contrast, the wasteland profiles exhibited irregular vertical SOC distribution, with anomalous enrichment observed at depths of 120 and 180 cm. This irregularity suggests that vertical variations in soil texture, particularly the strong adsorption capacity of clay layers, play a dominant role in SOC distribution. Furthermore, the SOC distribution in wasteland may reflect the region's original SOC state prior to human disturbance (Noppol *et al*., 2023).

Table 3. Summary of semivariance function parameters

Model	Nugget value	Sill value	Nugget effect (%)	Coefficient of determination	Residual
	C_o	C_o+C	$C_o/(C_o+C)$	R^2	RSS
Exponential	0.164	0.281	58.36	0.912	<0.01
Spherical	0.14	0.374	38.77	0.987	<0.01
Linear	0.226	0.574	39.37	0.924	<0.24
Gaussian	0.173	0.237	72.99	0.994	<0.01

In summary, the spatial heterogeneity of SOC in the study area is jointly controlled by anthropogenic factors (land management in cropland/orchard systems), natural vegetation patterns (woodland *vs*. grassland), and soil physicochemical properties (clay layers) (Wang Y.P. *et al*., 2024).

3.3. Characteristics of organic carbon spatial distribution

The spatial structural characteristics of SOC were analyzed by fitting theoretical variogram models to the SOC content variance function. Among the tested models (Exponential, Spherical, Linear, and Gaussian), the Gaussian model demonstrated the best performance, with the lowest residual value and the highest coefficient of determination ($R^2 = 0.994$) (Table 3), confirming its suitability for describing SOC spatial patterns in the oasis soils at the eastern edge of the Tarim Basin.

The nugget value (0.173) derived from the Gaussian model reflects microscale variability influenced by random factors, such as sampling spacing inconsistencies and measurement errors. A nugget effect of 72.99% indicates moderate spatial correlation at very small scales (near-zero sampling distances), suggesting that localized stochastic processes, rather than systematic spatial gradients, dominate SOC heterogeneity within the study area. This highlights the importance of high-resolution sampling to capture fine-scale SOC dynamics in similar arid ecosystems.

The spatial distribution of SOC in the study area was mapped using ordinary kriging interpolation based on the Gaussian model-derived semi-variogram, revealing a distinct "high-north-south, low-central" pattern characterized by fragmented high-value zones and ring-like scattering of intermediate values (Fig. 3). The central region is dominated by low SOC content (< 5 g kg^{-1}), covering 60.56% of the total area, likely influenced by intensive land use, erosion, or limited organic inputs. Intermediate SOC levels (5-8 g kg^{-1}) occupy 26.55% of the area, forming discontinuous transitional zones around the central core, while scattered high-SOC enclaves (> 8 g kg^{-1}) in the northwest and southeast (12.88% of the area) suggest localized hotspots potentially linked to wetland retention, vegetation density, or historical management practices. The absence of extensive contiguous high-SOC areas highlights the fragmented influence of both natural and anthropogenic drivers in this arid oasis ecosystem, consistent with the previously observed moderate small-scale spatial correlation (72.99% nugget effect). This spatial configuration underscores the dominance of stochastic processes, such as uneven irrigation or microtopographic variations, in shaping SOC heterogeneity at the eastern margin of the Tarim Basin.

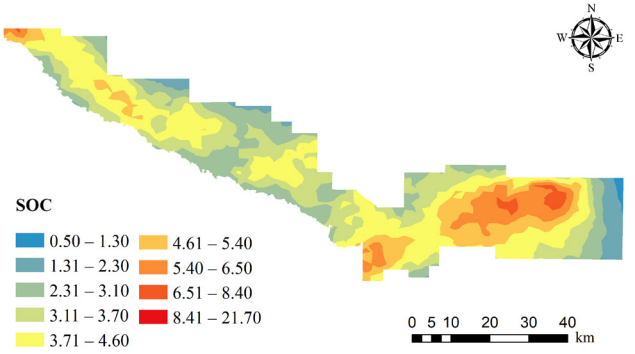

Fig. 3. Geospatial variability of soil organic carbon (SOC, g kg^{-1}) across the delineated region.

3.4. Identification of the main control factors

3.4.1. Data sources

The study area, located on the northeastern margin of the Tarim Basin, is characterized by fragmented farmland distribution due to urbanization pressures (Liu *et al*., 2024b) and serves as a critical zone for analyzing arid oasis agroecosystems. Influencing factors were selected based on representativeness, accessibility, and alignment with regional agricultural dynamics, categorized intonexternal environmental stochastic factors (climate variability, land use patterns) and inherent soil structure attributes (texture, chemical composition). The data collection integrated field surveys and existing databases, encompassing seven subcategories: chemical properties (TN, TP), physical characteristics (BD, porosity), soil texture types, soil classification, land use history, climatic variables (precipitation, temperature), and topographic features (elevation, slope). Over 20 specific indicators were systematically compiled to capture multi-dimensional drivers of SOC variation, with detailed metadata, including sources (laboratory

Table 4. Compilation of multisource data categories and provenance metadata

Index structure	Indicator classification	Indicator nam	Data sources	Variable type	Data handling
Soil structure endowment factors	Chemical feature	TN, TP, TK, TS, pH	Real measurement	continuous variable	statistical analysis
	Physical feature	BD	Real measurement	continuous variable	statistical analysis
	Texture type	clay, silt, clay	Real measurement	categorical variable	statistical analysis
External environmental random factors	Soil type	–	HWSD V2.0	categorical variable	extract to point
	Land use	plow land, garden land, forest land, uncultivated land	Survei Geologi China GeoCloud 3.0	categorical variable	extract to point
	Climate	average temperature, average rainfall	Pusat Data Sains Gurun Tundra Glasial Nasional	continuous variable	extract to point
	Terrain	surface roughness, topographic relief	Pusat Data Sains Gurun Tundra Glasial Nasional	continuous variable	raster calculation

BD – dulk density, HWSD – harmonized world soil database. Other explanations as in Table 2.

analyses, remote sensing, historical records) and spatial resolutions, summarized in Table 4. This comprehensive dataset bridges microscale soil processes and macroscale environmental gradients, supporting the interpretation of the observed fragmented SOC patterns and their linkage to anthropogenic activities (irrigation practices) and natural heterogeneity in this ecologically transitional region.

3.4.2. Accuracy comparison

Four regression models: OLS, RF, SVM, and BP were evaluated for predicting the SOC content in the arid oasis ecosystem of the Tarim Basin's eastern margin. As shown in Fig. 4, the RF model exhibited superior performance, achieving the highest coefficient of determination (R^2 = 0.7943) and the lowest prediction errors, with RMSE = 0.0180 and MAE = 0.0137. In contrast, the linear OLS model demonstrated limited efficacy, yielding the lowest accuracy (R^2 = 0.5714) and the highest errors (RMSE = 0.0229, MAE = 0.0183), while the SVM (R^2 = 0.7312) and BP models displayed intermediate performance. These results highlight the dominance of nonlinear machine learning approaches, particularly RF, in capturing complex interactions between SOC and environmental variables (climate gradients, soil texture heterogeneity) in fragmented land use conditions. The RF model's robustness aligns with its capacity to handle high-dimensional, nonparametric relationships and mitigate overfitting, as evidenced by its enhanced interpretability of environmental drivers (Zhang *et al*., 2023). The marked disparity between linear (OLS) and nonlinear (RF, SVM, BP) models underscores the critical role of algorithmic flexibility in addressing spatial heterogeneity and stochastic processes, such as uneven irrigation or microtopographic variability, which govern SOC distribution in this ecologically transitional region. This accuracy comparison reinforces the suitability of RF for spatially explicit SOC mapping in arid oasis agroecosystems characterized by fragmented environmental gradients.

3.4.3. Analysis of key factors

To elucidate the drivers of SOC spatial heterogeneity in the arid oasis ecosystem of the Tarim Basin's eastern margin, the RF model was employed to quantify the relative importance of climatic, topographic, pedogenic, and anthropogenic variables. As depicted in Fig. 5, soil chemical properties (TN, TP) emerged as the dominant predictor, explaining 29.3% of SOC variability, underscoring their pivotal role in carbon sequestration dynamics through nutrient cycling and organic-mineral interactions. Physical properties (BD, porosity: 18.2%) and land use type (14.2%) ranked second and third, respectively, reflecting the interplay between soil structure, agricultural management practices (tillage, irrigation), and SOC stabilization. Soil type (10.9%) further highlighted the influence of pedogenic processes unique to the region's saline-alkali and alluvial soils. In contrast, textural class (sand/clay fractions), climatic variables (aridity index), and topographic features (slope: 1.5%) exhibited marginal contributions, suggesting that the limited moisture availability and subdued relief in this low-energy landscape attenuate their direct impacts. Cumulatively, the soil physicochemical attributes accounted for 47.5% of SOC spatial variability, dominating over the extrinsic environmental factors. This hierarchy aligns with the study area's fragmented agroecosystem, where localized anthropogenic interventions (fertilization, crop rotation) and inherent soil heterogeneity override broader climatic or geomorphic gradients. The minimal topographic

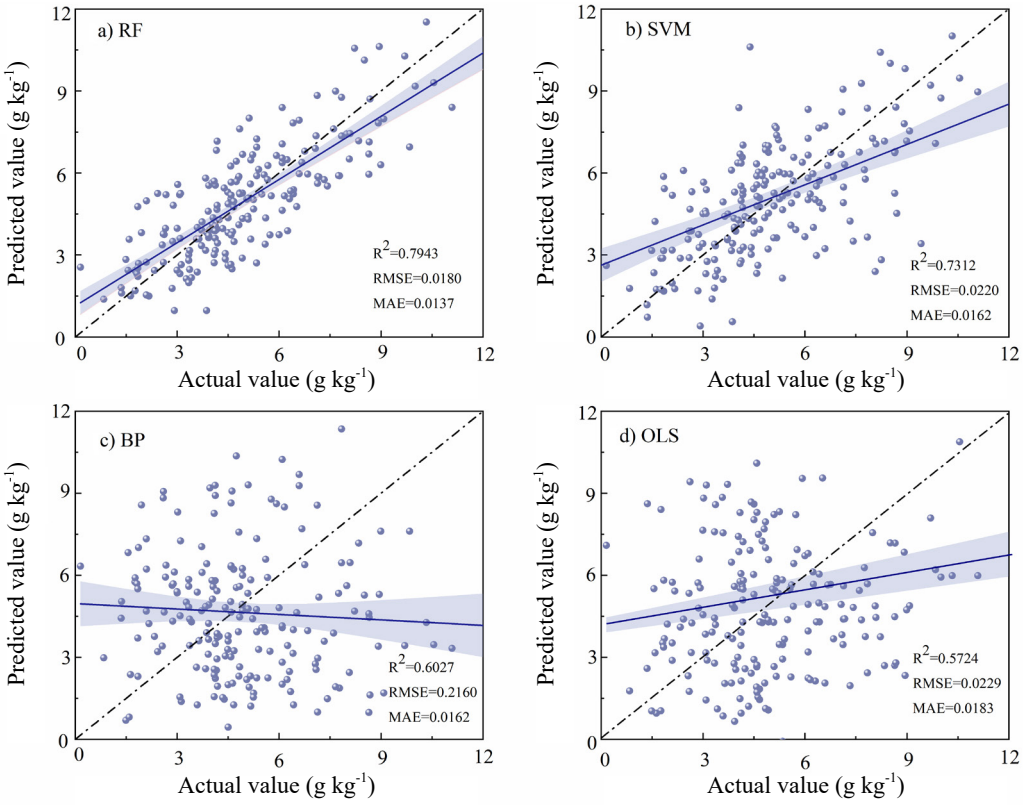

Fig. 4. Cross-model validation of machine learning predictions: observed *vs*. estimated value dispersion: a) RF – random forest, b) SVM – support vector machines, c) BP – back propagation, d) OLS – ordinary least squares.

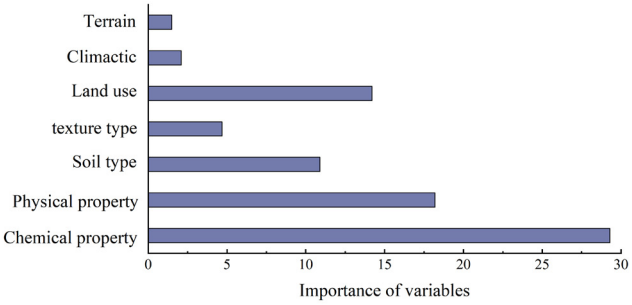

Fig. 5. Machine learning-driven quantification of environmental determinants on SOC pool variability.

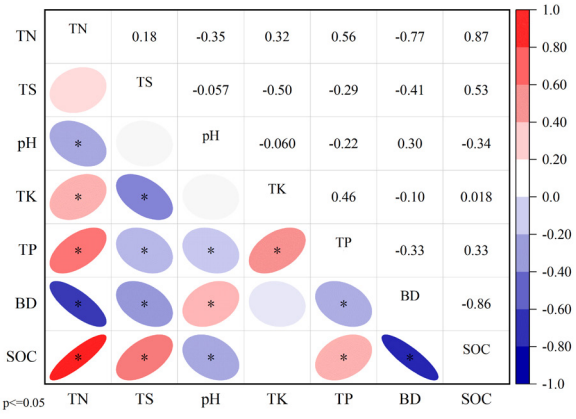

Fig. 6. Multivariate correlation heatmap of soil organic carbon (SOC) and Key Edaphic Drivers. The color gradient ranges from blue (r = -1.0) to red (r = +1.0), representing the continuum from perfect negative to positive correlation. TK – total potassium, TN – total nitrogen, TP – total phosphorus, TS – total salinity, pH – potential of hydrogen, BD – dulk density, SOC – soil organic carbon.

influence contrasts with humid regions, emphasizing the unique biogeochemical drivers governing SOC dynamics in arid transitional zones.

4. DISCUSSION

4.1. Mechanism of soil physical and chemical properties on SOC

The dynamics of SOC are intricately linked to soil physico-chemical properties, mediated through biogeochemical and structural interactions (Fig. 6). Nitrogen exhibits a strong positive correlation with SOC (r = 0.87), reflecting its critical role in organic matter stabilization through microbial biomass turnover and formation of organo-mineral complexes. In contrast, phosphorus shows negligible association with SOC (r = 0.018), likely due to its reduced bioavailability in the study area's alkaline soil (mean pH = 8.45), where high pH promotes phosphorus fixation via adsorption to calcium carbonate, decoupling it from SOC dynamics. Potassium, though essential for crop growth, is not explicitly correlated here but interacts indirectly through agricultural practices, such as fertilizer application (Fartyal *et al*., 2025).

Agricultural activities, including fertilization, crop rotation, and tillage, significantly influence nutrient availability and SOC stability (Dipakama *et al*., 2024). Moderate fertilizer use enhances soil microenvironment regulation, promoting organic carbon fixation, while excessive application accelerates SOC decomposition through acidification and microbial dysregulation. The strongly alkaline soil (pH=8.45) in the study region inhibits microbial diversity and enzyme activity (cellulase), thereby impeding cellulose decomposition in plant residues. This leads to short-term accumulation of particulate organic carbon (POC) but hinders its transformation into stable carbon pools. Additionally, the highly alkaline environment suppresses vegetation growth, reducing carbon input. Although decomposition rates are slowed, the decline in carbon input is more pronounced, ultimately resulting in a net decrease in SOC content and reduced stability (Yang *et al*., 2024). Alkaline conditions also inhibit surface vegetation growth, lowering organic carbon inputs from plant residues and humus migration, thereby reinforcing the negative correlation between pH and SOC (r = -0.34).

Total salt content demonstrates a moderate positive correlation with SOC (r = 0.53), as salinity in arid environments slows microbial decomposition, preserving SOC (Iseas *et al*., 2025; Gupta *et al*., 2024). Soil bulk density, however, emerges as a dominant physical regulator, showing a strong inverse relationship with SOC (r = -0.86). Low bulk density soils (< 1.4 g cm^{-3}) foster complex pore networks that physically protect SOC within aggregates and enhance microbial activity, facilitating organic matter decomposition and carbon enrichment. Conversely, high bulk density (> 1.6 g cm^{-3}) compacts soil structure, reducing pore connectivity and microbial habitat diversity. This compaction restricts root penetration and organic matter incorporation while promoting anaerobic conditions that slow residue decomposition but limit SOC retention (Tian *et al*., 2024).

These interactions underscore the dual role of agricultural management: practices like reduced tillage and balanced fertilization can mitigate bulk density increases and pH extremes, thereby enhancing SOC sequestration (Shen *et al*., 2024). Conversely, mismanagement exacerbates physicochemical constraints, accelerating SOC loss. The interplay of these factors highlights the sensitivity of SOC to both nutrient cycling and soil structural integrity in agroecosystems (Fenton *et al*., 2024).

4.2. Effect of land use on SOC

The study revealed significant variations in the soil organic carbon (SOC) content across the different land use types within the research area. Arable land exhibited the highest average SOC content (5.27 g kg^{-1}), while wasteland showed the lowest level (4.04 g kg^{-1}), with the following overall SOC ranking across five land use types: arable land > garden land > forest land > grassland > wasteland. The analysis of Fig. 7 demonstrated that garden land displayed the highest SOC peak value and the narrowest distribution range, indicating highly concentrated and uniform SOC distribution, likely attributable to intensive management practices, such as regular fertilization and irrigation. In contrast, arable land exhibited a comparably high SOC peak but a broader distribution range, reflecting substantial spatial heterogeneity in the SOC content among cultivated plots. This variability could be explained by differences in tillage intensity, crop rotation systems, and localized management practices. Nevertheless, the overall SOC levels in arable land remained significantly elevated compared to the other land use types, primarily due to prolonged anthropogenic interventions (Parajuli *et al*., 2024). Agricultural activities, including irrigation, fertilization, and plowing, have substantially enhanced soil physicochemical properties (bulk density, porosity, and water retention capacity), thereby promoting SOC accumulation. Conversely, wastelands experienced exacerbated SOC depletion under environmental stressors, such as arid climatic conditions, limited precipitation, intense surface evapotranspiration, frequent wind-sand activities, and sparse vegetation cover, which collectively reduced organic matter input and accelerated carbon loss (Ben *et al*., 2024).

The research further emphasized that land use practices directly and indirectly regulate SOC dynamics by modifying surface cover characteristics (vegetation types, litter accumulation), root distribution patterns, and soil properties (texture, moisture content, and nutrient composition), as demonstrated in previous studies (Ma Z.M. *et al*., 2024; Paramesha *et al*., 2024). For instance, arable and garden

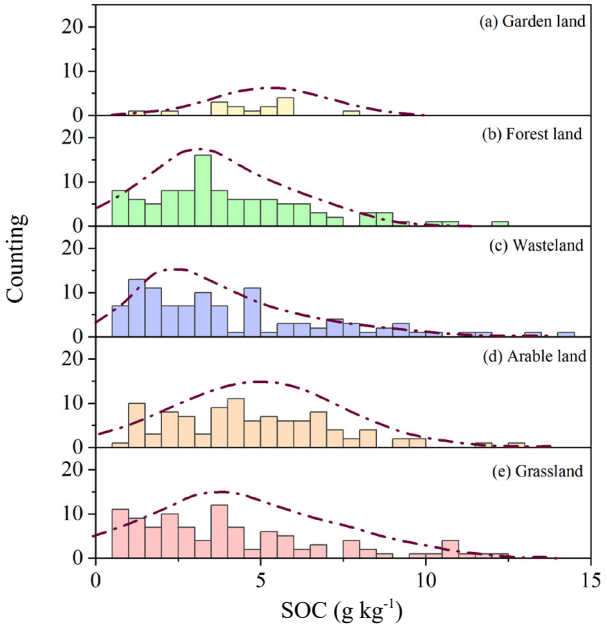

Fig. 7. Variations in SOC under contrasting land management practices.

lands benefited from human-selected fertile soil substrates and sustained water-fertilizer management, creating superior hydrothermal and nutrient conditions compared to natural or semi-natural ecosystems. These anthropogenic modifications established microenvironments conducive to SOC stabilization (Petersson *et al*., 2025). The findings highlight the dual role of human activities in arid region soil carbon management: rational agricultural practices may enhance carbon sequestration potential, whereas excessive disturbance or environmental stressors can amplify carbon loss risks. This underscores the necessity for balanced land management strategies that optimize SOC retention while mitigating ecosystem degradation in fragile environments (Joshi *et al*., 2025).

4.3. Mechanism of soil type on SOC

According to the World Soil Database, the main soil types in the study area include meadow soil, fluvo-aquic soil, aeolian sandy soil, irrigation-silted soil, and saline-alkali soil. The analysis has revealed that, among these five soil types, fluvo-aquic soil exhibits the highest SOC content (5.27 g kg^{-1}), while aeolian sandy soil shows the lowest level (4.29 g kg^{-1}). The SOC content follows this order: alluvial soil (5.27 g kg^{-1}) > irrigated desert soil (5.09 g kg^{-1}) > saline-alkali soil (4.52 g kg^{-1}) > meadow soil (4.35 g kg^{-1}) > aeolian sandy soil (4.29 g kg^{-1}), with statistically significant differences ($p<0.05$) observed between the soil types. This stratification reflects distinct pedogenic processes and land-use impacts across the soil continuum.

As illustrated in Fig. 8, aeolian sandy soil displayed a low SOC distribution peak and a relatively narrow range, indicating limited spatial variability. Its inherently low SOC levels likely stem from the minimal anthropogenic development, sparse vegetation cover, and insufficient organic matter input from plant litter. Furthermore, the region's arid climate, characterized by low precipitation and high evapotranspiration, coupled with suppressed microbial activity, restricts organic carbon replenishment (Zong *et al*., 2025). The coarse texture of aeolian sandy soil, dominated by sand particles, further impedes organic carbon stabilization and accumulation (Li C. *et al*., 2024).

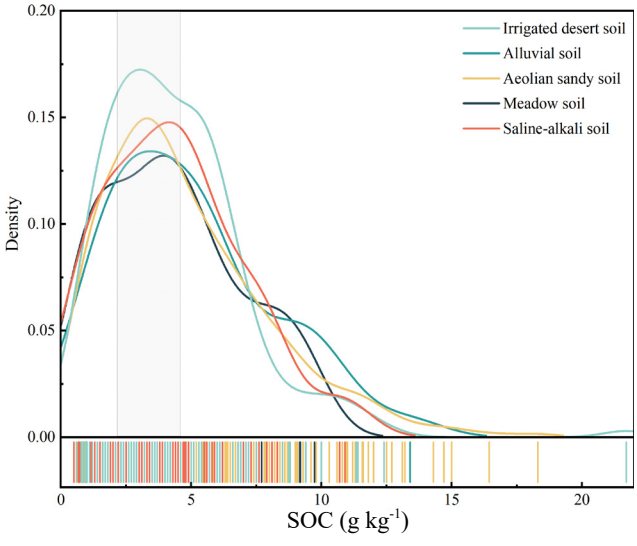

Fig. 8. Distribution patterns of SOC in different pedological classes.

In contrast, the alluvial soil demonstrated superior SOC retention, attributed to its favorable physicochemical properties (Borah *et al*., 2024). The lighter texture and higher clay content enhance soil structure, moisture retention, and biological activity, facilitating humus formation and organic carbon sequestration. Additionally, continuous cultivation practices, including sustained organic fertilizer application, improve nutrient availability and stimulate microbial decomposition, thereby promoting SOC accumulation (Li *et al*., 2024d). These anthropogenic interventions, combined with inherent soil characteristics, such as optimal bulk density and water-holding capacity, create a microenvironment conducive to organic carbon stabilization.

The findings underscore the critical role of soil type in regulating SOC dynamics, mediated by interactions between intrinsic soil properties (texture, clay content) and extrinsic factors, such as land management practices and climatic conditions. This highlights the necessity of soil-specific strategies for enhancing carbon sequestration in arid and semi-arid ecosystems (Oukhattar *et al*., 2025).

5. CONCLUSIONS

This study systematically investigated the spatial variability and driving mechanisms of soil organic carbon (SOC) in the oasis ecosystem at the eastern edge of the Tarim Basin. The key findings are summarized as follows:

1. The SOC content across the study area ranged from 0.50 to 21.70 g kg^{-1}, with a mean value of 4.47 g kg^{-1} and a coefficient of variation (CV) of 65.5%, indicating pronounced spatial heterogeneity. Spatially, SOC exhibited a distinct zonation pattern, characterized by higher concentrations in the northern and southern regions and lower values in the central zone. This distribution manifested as concentric annular bands interspersed with scattered hotspots, likely reflecting localized variations in land management practices and soil-forming processes.

2. The comparative evaluation of four predictive models: ordinary least squares (OLS), random forest (RF), support vector machines (SVM), and back propagation (BP), revealed that the RF model outperformed the others in capturing nonlinear relationships between SOC and environmental covariates. The RF algorithm demonstrated superior interpretability for variable importance (soil texture, land use type) and achieved the highest prediction accuracy (R^2 = 0.7943, RMSE = 0.0180), highlighting its robustness for SOC mapping in complex arid ecosystems.

3. The spatial heterogeneity of SOC was governed by synergistic interactions between stochastic external factors (anthropogenic interventions, wind-sand disturbances) and intrinsic soil properties (texture, bulk density). The geostatistical analysis confirmed significant spatial autocorrelation (Moran's $I > 0.4$, $p < 0.01$) among the influencing variables. Among the evaluated drivers, soil physicochemical properties (clay content, moisture retention), land use type, and soil type collectively explained >70% of SOC variability.

4. These findings underscore the critical role of anthropogenic activities in modulating SOC dynamics within arid oasis ecosystems. The RF-based spatial prediction framework provides a reliable tool for identifying SOC sequestration hotspots and guiding precision land management. To mitigate carbon loss risks, priority should be given to optimizing irrigation regimes, promoting organic amendments in sandy soils, and maintaining vegetation cover in marginal lands. Future studies should integrate long-term monitoring to assess climate feedbacks on SOC stability in these fragile agroecosystems.

6. ACKNOWLEDGMENTS

We thank to the Key Laboratory of Coupling Process and Effect of Natural Resources Elements, Geological Survey Projects of the China Geological Survey, and Science and Technology Innovation Fund for the Integrated Natural Resources Survey Command Centre. We are equally grateful to the journal editors and reviewers for their support in the publication of the article.

Conflict of interest. The authors declare that they have no conflict of interest.

Data availability. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

7. REFERENCES

Abdellafou, B.K., Zidi, K., Aljuhani, A., Taouali, O., Harkat, M.F., 2025. Machine learning for smart soil monitoring. Comput. Mater. Con. 83, 3007-3023. https://doi.org/10.32604/CMC.2025.063146

Annalisa, S., Zhang, Y.K., Huang, J.Y., Hu, J., Keith, P., Alfred, H.E., 2024. Rates of soil organic carbon change in cultivated and afforested sandy soils. Agric. Ecosys. Environ. 360, 109632. https://doi.org/10.1016/J.AGEE.2023.108785

Batista, V.P., Möller, M., Schmidt, K., Waldau, T., Seufferheld, K., Htitiou, A., *et al*., 2025. Soil-erosion events on arable land are nowcast by machine learning. Catena 256, 109080. https://doi.org/10.1016/J.CATENA.2025.109080

Ben A.A., A., Shukla, M., 2024. Assessment of soil organic and inorganic carbon stocks in arid and semi-arid rangelands of southeastern New Mexico. Ecol. Indic. 166, 112398-112398. https://doi.org/10.1016/J.ECOLIND.2024.112398

Boubehziz, S., Piccini, C., González, M.A.J., Almendros, G., 2024. Spatial distribution of soil organic carbon quality descriptors determining factors that affect its sequestration in Northeast Algeria. J. Environ. Manag. 358, 120772. https://doi.org/10.1016/J.JENVMAN.2024.120772

Borah, B., Parmar, P., 2024. Soil organic carbon dynamics: drivers of climate change-induced soil organic carbon loss at various ecosystems. Int. J. Environ. Climate Change 14, 153-174. https://doi.org/10.9734/IJECC/2024/V14I104477

Cao, X.M., Cui, M.C., Xi, L., Feng, Y.M., 2024. Spatial-temporal process of land use/land cover and desertification in the Circum-Tarim Basin during 1990-2020. Land. 13, 735. https://doi.org/10.3390/LAND13060735

Dipakama, C.M., Dieu, N.J., Watha, N.N., Nguelet, M.I., Kimpouni, V., 2024. Impact of agricultural activities on soil particle size, ph, and organic matter in the Dimonika Biosphere Reserve, Mayombe (Republic of Congo). Environ Pollut. 14, 1-1. https://doi.org/10.5539/EP.V14N1P1

Fartyal, A., Bargali, S.S., Bargali, K., Negi, B., 2025. Changes in soil properties, organic carbon, and nutrient stocks after land‐use change from forests to grasslands in Kumaun Himalaya, India. Land Degrad. Dev. 36, 2438-2457. https://doi.org/10.1002/LDR.5507

Fenton, O., Bondi, G., Bracken, C.J., O'Sullivan, L., Sangil, L.L., Tuohy, P., *et al*., 2024. Relative and absolute difference in soil organic carbon stocks in grassland soils in Ireland: Impact of rock fragments, bulk density and calculation methods. Geoderma Reg. 36, e769. https://doi.org/10.1016/J.GEODRS.2024.E00769

Fu, P., Clanton, C., Demuth, M.K., Goodman, V., Griffith, L., Young, K.M., *et al*., 2024. Accurate quantification of 0-30 cm soil organic carbon in croplands over the continental United States using machine learning. Remote Sensing 16, 2217-2217. https://doi.org/10.3390/RS16122217

Gupta, V., Bangar, K.S., Parmar, B.B., Bangre, J., Kumar, K., Yadav, B., 2024. Unlocking the secrets of soilhealth: exploring the influence of natural and synthetic nutrients on soil organic carbon and active and passive carbon pool in vertisols. Int. J. Plant Sci. 36, 556-561. https://doi.org/10.9734/IJPSS/2024/V36I64658

Huong, M.T.N., Quynh, P.T.L., Van, V.H., Tu, C.V., 2022. Biodegradable and seasonal variation of organic carbon affected by anthropogenic activity: a case in Xuan Thuy Mangrove Forest, North Vietnam. Water 14, 773-773. https://doi.org/10.3390/W14050773

Huyzentruyt, M., Belliard, J.P., Saintilan, N., Temmerman, S., 2024. Identifying drivers of global spatial variability in organic carbon sequestration in tidal marsh sediments. Sci. Total Environ. 957, 177746. https://doi.org/10.1016/J.SCITOTENV.2024.177746

Iseas, S.M., Sainato, M.C., Romay, C., 2025. Supplemental irrigation in the humid Pampean region: Effects on soil salinity, physical properties, nutrients and organic carbon. Soil Till. Res. 248, 106421. https://doi.org/10.1016/J.STILL.2024.106421

Joshi, K.R., Garkoti, C.S., 2025. Ecosystem carbon storage, allocation and carbon credit values of major forest types in the central Himalaya. Carbon Res. 4, 7-7. https://doi.org/10.1007/S44246-024-00159-4

Li, C., Ran, M., Song, L., 2024. Temperature effects on cropland soil particulate and mineral-associated organic carbon are governed by agricultural land-use types. Geoderma 448, 116942. https://doi.org/10.1016/j.geoderma.2024.116942

Li, S., Manuel, D., Ding, J.X., *et al.,* 2024. Intrinsic microbial temperature sensitivity and soil organic carbon decomposition in response to climate change. Global Change Biol. 30, e17395. https://doi.org/10.1111/GCB.17395

Li, W.D., Li, A.W., Cheng, J.L., Chen, D., Mao, Y.R., Chen, X.Y., *et al.,* 2024. The impact of agricultural land use patterns and soil types on the components of soil organic carbon in the plow layer in the Tuojiang River Basin. J. Ecol. Rural Environ. 3, 1-15. https://doi.org/10.19741/j.issn.1673-4831.2024.0533

Li, X.L., Zhang, Q.W., Li, M.N., Shi, Y.L., Yu, B.W., Jing, X.K., *et al.,* 2024. The interaction of black soil long gentle slope topography and horizontal ridges on the spatial differentiation of soil organic carbon. Trans. Chin. Soc. Agric. Eng. 40, 103-113. https://doi.org/10.11975/j.issn.1002-6819.202310175

Liu, G.L., Yin, G., Kurban, A., Aishan, T., You, H.L., 2016. Spatiotemporal dynamics of land cover and their impacts on potential dust source regions in the Tarim Basin, NW China. Environ. Earth Sci. 75, 1471-1477. https://doi.org/10.1007/s12665-016-6269-y

Liu, J.Y., Jiang, L., Yin, L.H., Hu, H.L., Wang, L.L., Sun, J.G., 2023. Analysis of soil nutrient characteristics and main controlling factors in the oasis area of the Northeast Edge of the Tarim Basin. Northwest Geol. 56, 141-152. https://doi.org/10.12401/j.nwg.2023088

Liu, J.Y., Li, R.Y., Liang, Y.C., Liu, L., Yin, F., Tang, S., *et al.,* 2024a. Prediction of cadmium content in oasis soils on the eastern edge of the tarim basin based on feature selection and machine learning, and health risk assessment. Environ. Sci. 45, 4802-4811. https://doi.org/10.13227/j.hjkx.202308010

Liu, J.Y., Yin, F., Liu, L., 2024b. Ecological stoichiometric characteristics and driving factors of soil in desert oasis areas. China Environ. Sci. 44, 300-309. https://doi.org/10.3969/j.issn.1000-6923.2024.01.031

Liu, D., Sun, J.Q., Yu, Z.B., 2024. Multi-layer soil moisture inversion based on machine learning. J. Hohai Univ. Nat. Sci. 52, 7-14. https://doi.org/10.3876/j.issn.1000-1980.2024.03.002

Ma, R., Li, X.Q., Tie, C.X., Hou, Q.Y., Liu, L.F., 2024. Distribution characteristics and influencing factors of soil organic carbon under different land use types in the Loess Hilly and Gully Region. J. Agric. Resour. Environ. 23, 1-12. https://doi.org/10.13254/j.jare.2024.0510

Ma, Z.M., Zhang, M.L., Liu, X.Y., 2024. Estimation of soil organic carbon content in Gannan Grassland Based on SSA-Optimized CatBoost. Environ. Sci. 21, 1-17. https://doi.org/10.13227/j.hjkx.202408081

Mackessy, F., McCarthy, E., Broderick, E., O'Donnell, B., Quille, P., 2024. Investigating the accuracy and comparability of various lime prediction methods for Irish grassland mineral soils. Soil Use Manag. 40, 1-12. https://doi.org/10.1111/SUM.13034

Neupane A., Lazicki P., Mayes M.A., Lee J.,·Jagadamma S., 2022. The use of stable carbon isotopes to decipher global change effects on soil organic carbon: present status, limitations, and future prospect. Biogeochemistry 160, 315-354. https://doi.org/10.1007/S10533-022-00963-3

Noppol, A., Sukanya, S., Praeploy, K., Monthira, Y., Ryusuke, H., 2023. Variations of soil properties and soil surface loss after fire in rotational shifting cultivation in Northern Thailand. Front Environ. Sci. 11, 1213181. https://doi.org/10.3389/fenvs.2023.1213181

Noppol, A., Sukanya, S., Ryusuke, H., 2021. Impact of burning on soil organic carbon of maize-upland rice system in Mae Chaem Basin of Northern Thailand. Geoderma 392, 115002. https://doi.org/10.1016/j.geoderma.2021.115002

Oukhattar, M., Gadal, S., Robert, Y., Saby, N., Houmma, H.I., Keller C., 2025. Variability analysis of soil organic carbon content across land use types and its digital mapping using machine learning and deep learning algorithms. Environ. Monit. Assess. 197, 535-535. https://doi.org/10.1007/S10661-025-13972-0

Petersson, T., Antoniella, G., Perugini, L., Chiriacò, V.M., Chiti, T., 2025. Carbon farming practices for European cropland: A review on the effect on soil organic carbon. Soil Till. Res. 247, 106353-106353. https://doi.org/10.1016/J.STILL.2024.106353

Parajuli, B., Lamichhane, N., Monokrousos, N., Pokhrel, P.C., Yadav, P.K.R., 2024. Influence of land-use practices on soil organic carbon and microbial biomass in coffee and orange agroecosystems. Land 13, 2076-2076. https://doi.org/10.3390/LAND13122076

Paramesha, V., Kumar, P., Francaviglia, R., Nath, A.J., Mishra, G., Kumar, R.M., *et al.,* 2024. Evaluating land use and climate change effects on soil organic carbon. A simulation study in coconut and pineapple systems in west coast India. Catena 248, 108587. https://doi.org/10.1016/J.CATENA.2024.108587

Qin, Z.L., Yang, X.M., Song, Z.L., Wu, S.C., Fang, X.H., Peng, B., 2020. The Impact of Parent Material and Land Use on the Chemical Composition of Soil Organic Carbon. Chin. J. Soil Sci. 51, 621-629. https://doi.org/10.19336/j.cnki.trtb.2020.03.16

Qin, D.R., Huang, Y.M., Huang, Q., Xu, F.J., Shen, J.K., 2024. Contribution of microbial residual carbon in different soil layers of typical forest-grassland on the loess plateau to soil organic carbon components and its influencing factors. Environ. Sci. 44, 1-16. https://doi.org/10.13227/j.hjkx.202408038

Sanleandro, P.M., García, A.M.G., Bernardeau, A.B., Vázquez, J.M.G., Linares, M.A.A., 2023. Influence of the type and use of soil on the distribution of organic carbon and other soil properties in a sustainable and resilient agropolitan system. Forests 14, 2063. https://doi.org/10.3390/F14061085

Shen, Z., Han, T.F., Huang, J., Li, J.W., Daba, N.A., Gilbert, N., *et al.,* 2024. Soil organic carbon regulation by pH in acidic red soil subjected to long-term liming and straw incorporation. J. Environ. Manag. 367, 122063. https://doi.org/10.1016/J.JENVMAN.2024.122063

Spotorno, S., Gobin, A., Vanongeval, F., Borghi, A.D., Gallo, M., 2024. Carbon Farming practices assessment: Modelling spatial changes of soil organic carbon in Flanders, Belgium. Sci. Total Environ. 922, 171267. https://doi.org/10.1016/J.SCITOTENV.2024.171267

Szostek, M., Szpunar-Krok, E., Pawlak, R., Stanek-Tarkowska, J., Ilek, A., 2022. Effect of different tillage systems on soil organic carbon and enzymatic activity. Agronomy 12, 208-208. https://doi.org/10.3390/agronomy12010208

Tian, L.J., Shao, G.C., Gao, Y., Song, E.Z., Lu, J., 2024. Effects of biochar on soil organic carbon in relation to soil nutrient

contents, climate zones and cropping systems: A Chinese meta-analysis. Land 13, 1608. https://doi.org/10.3390/LAND13101608

Wang, L., Fan, H.B., Wu, Z.H., Lv, T.G., Tan, Y.Z., 2024. Spatial patterns and influencing factors of organic carbon content in cultivated layer soils in the Eastern Sichuan Hilly area. Trans. Chin. Soc. Agric. Eng. 40, 169-178. https://doi.org/11.2047.S.20241128.2048.090

Wang, Y.P., Li, R.F., Yan, W.M., Han, X.Y., Liu, W.Z., Li, Z., 2024. Variations in soil organic carbon after farmland conversion to apple orchard. Agronomy 14, 963. https://doi.org/10.3390/agronomy14050963

Wang, Z., Kumar, J., Leff, S.R.W., Todd-Brown, K., Mishra, U., Sihi, D., 2024. Upscaling soil organic carbon measurements at the continental scale using multivariate clustering analysis and machine learning. J. Geophys. Res. Biogeosci. 129,e2024JG018359.https://doi.org/10.1029/2023JG007702

Xia, S., Song, Z.L., Yu, B.B., Fan, Y.R., Tony, V., Guo, L.D., *et al.,* 2024. Land use changes and edaphic properties control contents and isotopic compositions of soil organic carbon and nitrogen in wetlands. Catena 241, 108031. https://doi.org/10.1016/J.CATENA.2024.108031

Xiong, K., Jiang, X.Y., Huang, S.Q., Guan, S., Zou, X.B., Chen, C.T., *et al.,* 2024. Variations in iron-bound organic carbon in soils along an altitude gradient and influencing factors in a subtropical mountain ecosystem of southern China. J. Soil Sediment. 24, 3180-3194. https://doi.org/10.1007/S11368-024-03873-Z

Yang, Q., Qu, K.M., Yang, S., Sun, Y., Zhang, Y., Zhou, M.Y., 2021. Environmental factors affecting regional differences and decadal variations in the buried flux of marine organic carbon in eastern shelf sea areas of China. Acta Oceanol. Sin. 40, 26-34. https://doi.org/10.1007/S13131-020-1601-5

Yang, X., Bao, Y.W., Li, B.W., Wang, R.X., Sun, C., Ma, D.H., *et al.*, 2024. Effects of fertilization applications on soil aggregate organic carbon content and assessment of their influencing factors: A meta-analysis. Catena 242, 108135. https://doi.org/10.1016/J.CATENA.2024.108135

Yin, J., Xue, S.S., Xu, X.T., Yang, G.S., Zhang, W., Ma, S.Z., 2012. Research on the changes in land use landscape patterns in the agricultural reclamation area of the Tarim Basin, Xinjiang over the Past 35 Years. Chinese J. Soil Water Conservation 7, 54-56. https://doi.org/10.14123/j.cnki.swcc.2012.07.028

Zhang, J.H., Zhu, L.Q., Li, G.D., Zhao, F., Qin, J.T., 2021. Spatial characteristics of soil carbon and nitrogen in the transitional zone between Northern and Southern China and the boundary between warm temperate and subtropical zones. Acta Geographica Sinica 76, 2269-2282. https://doi.org/10.11821/dlxb202109016

Zhang, X.S., He, B., Sabri, S.M.M., Mohammed, A., Vladimirovich, U.D., 2022. Soil liquefaction prediction based on bayesian optimization and support vector machines. Sustainability 14, 11944. https://doi.org/11944.10.3390/SU141911944

Zhang, H.T., Wang, J.H., Zhang, Y.C., Qian, H.Y., Xie, Z.Y., Hu, Y.F., *et al*., 2023. Soil organic carbon dynamics and influencing factors in the Zoige Alpine Wetland from the 1980s to 2020 based on a random forest model. Land 12, 1973. https://doi.org/10.3390/LAND12101923

Zong, M.M., Abalos, D., Chen, J., Liang, Z., Li, Y., Elsgaard, L., *et al*., 2025. Ten-year effects of perennial cropping systems on soil organic carbon stock and stability in sandy soils: Mechanisms and biochemical drivers. Eur. J. Agron. 168, 127639. https://doi.org/10.1016/J.EJA.2025.127639