Water and Soil Management and Modeling

Online ISSN: 2783 - 2546

University of Mohaghegh Ardabili

Soil texture mapping: a novel approach combining interpolation techniques and decision tree classifiers

Samir Boudibi*[1], Zineeddine Benguega[1,2], Haroun Fadlaoui[1], Zine-eddine Khomri[1], Azeddine Aissaoui[1] Bachir Sakaa[1], Tarik Otmane[1], Narimen Bouzidi[1]

[1] Centre de Recherche Scientifique et Technique sur les Régions Arides, CRSTRA, Biskra, Algeria.
[2] Kasdi Merbah University, Department of Agronomic Sciences, Laboratory of Saharan Bioresources: Preservation and Valorization Ouargla, Algeria.

Abstract

This study proposes a reproducible and GIS-based methodology for digital soil texture mapping by integrating geostatistical interpolation with deterministic decision tree classifiers (DTCs) derived from the United States Department of Agriculture (USDA) soil texture classification system. A total of 68 topsoil samples (0–20 cm) were collected across the irrigated area of northern Biskra province (southeastern Algeria) and analyzed for sand, silt, and clay contents. Among the most commonly applied interpolation techniques, ordinary kriging (OK), simple kriging (SK), and inverse distance weighting (IDW) were tested to generate continuous spatial distribution maps of soil particle fractions. Since the objective of this research was methodological demonstration rather than comprehensive benchmarking of interpolation algorithms, the method showing slightly better cross-validation (LOOCV) performance was selected. OK produced marginally lower RMSE values (15.93% for sand and 13.11% for silt) and satisfactory coefficients of determination (R^2 = 0.758 for sand and 0.687 for silt) and was therefore adopted. To preserve the compositional constraint (sand + silt + clay = 100%), clay content was derived from interpolated sand and silt maps. Four deterministic DTCs were implemented within the GIS environment to convert particle fraction rasters into continuous USDA texture classes. The final texture map demonstrated an almost perfect agreement with observed classifications (Kappa coefficient = 0.898). The proposed framework emphasizes methodological simplicity, transparency, and applicability under moderate sampling density without reliance on auxiliary environmental covariates or complex machine learning models. Although interpolation uncertainty may influence classification near texture boundaries, the approach provides a practical and scientifically robust solution for soil texture mapping in data-limited regions.

Keywords: Digital mapping, Decision tree classifier, GIS, Interpolation, Soil texture.

Article Type: Research Article
Academic Editor: Raoof Mostafazadeh

*Corresponding Author, E-mail: Samir.boudhibi@gmail.com; s.boudibi@mail-crstra.dz

Citation : Boudibi, S., Benguega, Z., Fadlaoui, H., Khomri, Z., Aissaoui, A., Sakaa, B., Otmane, T., Bouzidi, N. (2026). Soil texture mapping: a novel approach combining interpolation techniques and decision tree classifiers Water and Soil Management and Modelling, 6(2) (Special Issue: New Approaches to Water and Soil Management and Modeling), 287-302. doi: 10.22098/mmws.2026.19187.1767

Received: 20 December 2025, Received in revised form: 09 January 2026, Accepted: 23 January 2026, Published online: 03 June 2026

Water and Soil Management and Modeling, Year 2026, Vol. 6, No.2 (Special Issue), pp. 287-302

Publisher: University of Mohaghegh Ardabili

1. Introduction

Soil texture is a crucial natural attribute that plays a pivotal role in determining soil physical behavior and environmental functions (Song et al., 2023). The evaluation, analysis, and mapping of the spatial distribution and variability of soil texture are essential in agricultural regions due to their effects on plant development and yields (Liao et al., 2013; Seyedmohammadi et al., 2019). Soil texture is defined by the particle-size distribution of mineral soil fractions smaller than 2 mm, expressed as relative proportions (by mass) of clay (< 0.002 mm), silt (0.002–0.05 mm), and sand (0.05–2 mm) particles (Song et al., 2023). It is considered one of the most influential soil physical properties affecting crop growth and development (Yousif et al., 2023), influencing soil moisture content, water retention capacity, infiltration rate, drainage, fertilizer mobility, soil biodiversity, porosity, and susceptibility to erosion (Chau et al., 2011; Zhao et al., 2015; Bidkhani and Mobasheri, 2018; Ravikumar, 2022).

Soil texture classification is based on the proportion of the aforementioned particle-size fractions, which are determined in the laboratory using standard methods such as the pipette method. Each soil sample is classified according to the United States Department of Agriculture (USDA) soil textural classification system (Ravikumar, 2022). However, point-based laboratory determinations alone do not provide continuous spatial representation, which is necessary for land management and environmental modeling applications. Transitioning from discrete field measurements to spatially continuous representations is achieved through Digital Soil Mapping (DSM), which has emerged as a major sub-discipline of soil science (Taghizadeh-Mehrjardi et al., 2014; Minasny and McBratney, 2015). DSM is defined as the computer-assisted production of digital representations of soil types or properties using field and laboratory observations combined with spatial inference models (Lagacherie et al., 2007; Grunwald, 2009).

The most commonly applied DSM approaches include spatial interpolation techniques and machine learning (ML) algorithms (Santra et al., 2017; Qu et al., 2024). Spatial interpolation methods rely primarily on spatial autocorrelation structures derived from measured data, whereas ML-based approaches typically integrate environmental covariates, terrain attributes, and remote sensing data to model soil variability (De Caires et al., 2025). In recent years, increasing attention has been given to advanced ML algorithms and high-resolution satellite imagery for soil texture prediction (Wadoux et al., 2020; Yousif et al., 2023). While these approaches often improve predictive performance, they may require extensive datasets, significant computational resources, advanced technical expertise, and substantial preprocessing efforts. Furthermore, complex models may reduce interpretability and reproducibility, particularly when strict categorical systems such as the USDA texture triangle must be consistently applied at pixel level.

Spatial interpolation techniques remain widely used due to their conceptual simplicity and robustness. These techniques may be deterministic, such as inverse distance weighting (IDW), polynomial interpolation, and radial basis functions, or stochastic (geostatistical), such as kriging methods including ordinary kriging (OK), simple kriging (SK), universal kriging, indicator kriging, disjunctive kriging, and cokriging (Johnston et al., 2001; Ahmadi and Sedghamiz, 2007)). Many GIS platforms, including ArcGIS with its Geostatistical Analyst extension, provide tools for variogram modeling, cross-validation assessment, and raster prediction from limited sampling data (Karp et al., 2024). Among geostatistical approaches, ordinary kriging (OK) is one of the most frequently applied techniques in environmental studies due to its flexibility and ability to provide unbiased estimates under the assumption of an unknown but locally constant mean (Oliver and Webster, 2014; Bradaï et al., 2016; Bajjali, 2023). Variography involves fitting a theoretical semivariogram to the experimental semivariogram to quantify spatial dependence (Abzalov, 2016; Anthony et al., 2025). Cross-validation, typically implemented as a leave-one-out (LOOCV) procedure, evaluates model accuracy by successively removing individual observations and predicting their values from neighboring samples (Bernoux et al., 2007; Maroufpoor et al., 2017; ESRI, 2024, Wani et al., 2024). Decision trees are flowchart-like tree structures used across various fields to capture

descriptive decision-making knowledge from supplied data (Han et al., 2012, Hateffard et al., 2019). Classification trees are applied when the response variable is categorical, whereas regression trees are used for continuous outcomes (Szczerbicki, 2001; Wu et al., 2016; Rahmati et al., 2019). Despite the widespread application of decision trees in predictive modeling, limited attention has been given to implementing fully deterministic rule-based classifiers that strictly follow USDA soil texture conditions within a raster-based GIS environment.

In many arid and semi-arid regions, soil datasets are often sparse and auxiliary environmental variables may be limited or unavailable. Under such conditions, there is a need for transparent, computationally efficient, and reproducible methodologies capable of generating reliable soil texture maps using limited sampling data and widely accessible GIS tools. In this study, a simple and reproducible GIS-based framework is proposed for mapping soil texture classes using a sparse dataset of soil sampling points. The methodology consists of two main steps: (i) applying an appropriate interpolation technique to generate continuous spatial distribution maps of soil particle fractions (sand and silt), while preserving the compositional constraint for clay, and (ii) integrating deterministic decision tree classifiers derived from USDA soil texture rules to convert particle fraction rasters into a comprehensive soil texture classification map. The novelty of the proposed approach lies in combining interpolation outputs with fully deterministic USDA consistent classification rules at pixel level, without reliance on auxiliary environmental covariates or complex machine learning models.

2. Material and methods

2.1. Study area

The study area covers approximately 404.8 Km^2 ha and is situated between 34° 48’ to 35° 03’ N and 5° 09’ - 5° 45’ E, in the northern region of Biskra province, southeastern Algeria (Figure 1). It experiences an arid climate, with annual rainfall seldom exceeding 150 mm and mean temperatures ranging from 10.9°C in January to 35°C in July. The average annual relative humidity stands at 42%, while the mean annual evaporation reaches 2645 mm. In September 2017, the land surface temperature, calculated using satellite images reached45°C (Boudibi 2021). Based on the soil map of Algeria, produced by Durand and Barbut (1938) and further developed by Benchetrit (1956), the study area has fundamental alluvial characteristics encompassed by wind erosion soils. Slontchak-type saline soils are found in the Selga Saadoun synclinal depression. Aeolian accumulation soils, forming dunes, dominate the western part of the study area due to wind erosion (Boudibi 2021).

2.2. Sampling and laboratory analysis

In September 2017, a total of 68 soil samples are gathered from 68 sites, with their coordinates recorded via a Global Positioning System (GPS). These sampling sites are strategically selected to encompass the entirety of the irrigated area, primarily from farms. Subsequently, in the laboratory, the samples underwent air drying and are filtered through a 2 mm sieve. The soil particle fractions (clay, sand, and silt) are determined using the pipette method. Figure 2 illustrates the methodology flowchart employed in this study.

2.3. Interpolation technique

Interpolation techniques, whether geostatistical (e.g., kriging) or deterministic (e.g., IDW), aim to estimate a variable at unsampled locations and generate a continuous spatial representation of the study area. Among geostatistical approaches, kriging methods including SK, OK, universal kriging, and indicator kriging are widely recognized as robust interpolation techniques in environmental studies (Seyedmohammadi et al., 2019). In the present study, OK, SK, and IDW were evaluated using the Geostatistical Analyst extension in ArcGIS 10.8. Prior to interpolation, the normality of soil particle fractions was assessed using the Shapiro-Wilk test. Although moderate deviations from normality were observed, kriging methods have been shown to remain reliable under such departures, particularly when spatial dependence is well defined and predictive performance is validated through cross-validation (Armstrong and Boufassa, 1988; Webster and Oliver, 2007; Wani et al., 2024). Therefore, interpolation was conducted on the original data scale. Model performance

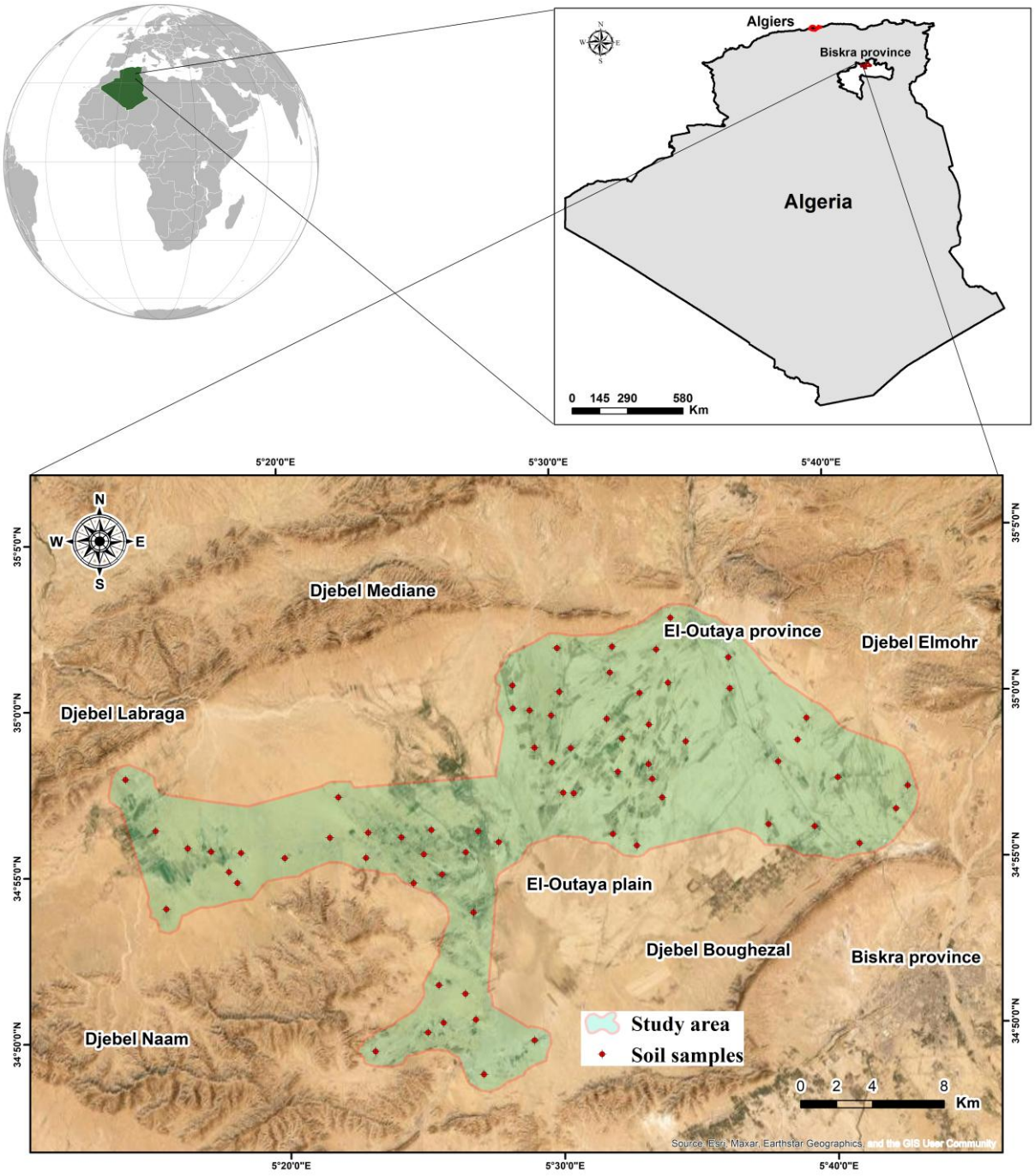

Figure 1. Study area and soil sampling points

was evaluated using automatic LOOCV. Based on the LOOCV statistics (ME, RMSE and R^2), OK demonstrated slightly superior predictive performance and was selected for subsequent spatial prediction of soil particle fractions. The primary tool in geostatistical techniques is the semivariogram γ(h), which quantifies the spatial dependence between neighboring observations (Webster and Oliver 2007; Arslan 2012; Şen, 2020). The equation for the semivariogram is typically expressed as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (1)$$

Here:

- γ(h) represents the semivariance at lag distance h.

- N(h) denotes the number of pairs of sample points separated by lag distance h.
- Z(xi) and Z(xi+h) are the values of the variable being analyzed at locations xi and xi+h, respectively.

Spherical models (Figure 3) are employed to fit the experimental semi variograms using VARIOWIN software v2.2. The selection of the best-fitting model is based on the indicative goodness of fit (IGF) provided by the software VARIOWIN. A good fit is indicated by an IGF value as close as possible to 0 (Pannatier 1996).

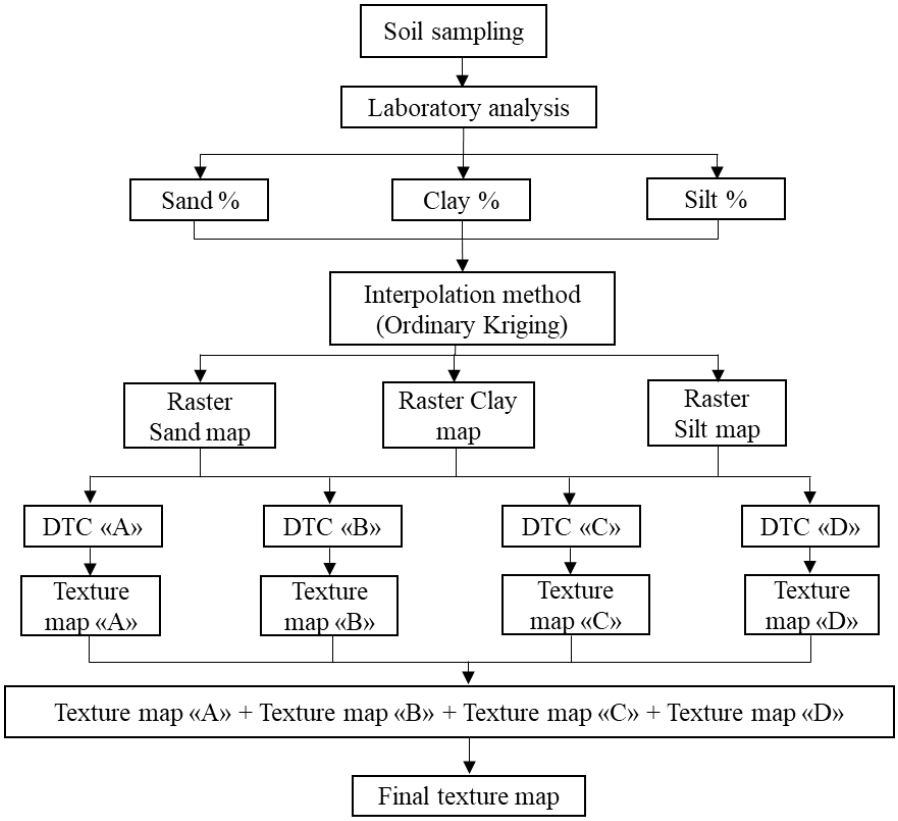

Figure 2. Flowchart of the developed methodology

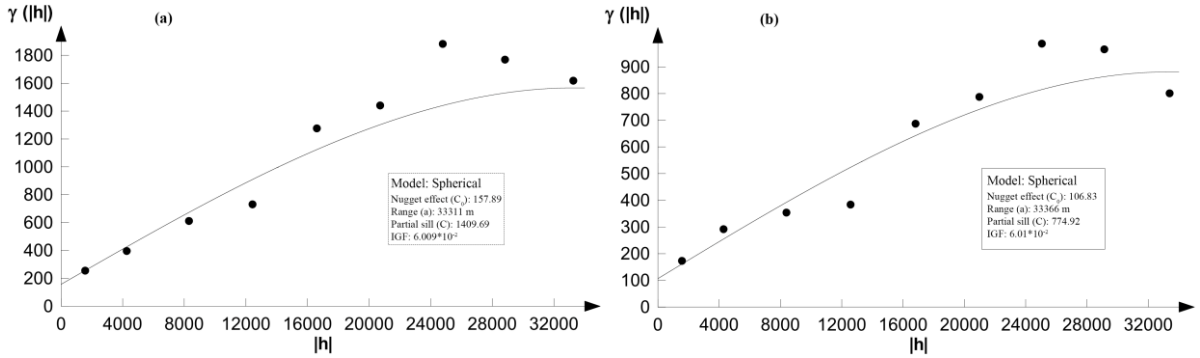

Figure 3. Omnidirectional semivariograms of the soil particle fractions: (a) silt semivariogram and (b) sand semivariogram

The primary parameters of the theoretical semivariograms (C, C_0, and a) generated using VARIOWIN software are utilized in the Geostatistical Analyst Extension of ArcGIS 10.8 software to generate various prediction maps for the sand and silt fractions (Figure 4).

The prediction map of the clay (Figure 4c) is obtained using Raster calculator in ArcGis software, following the equation:

$$Raster\ clay\ map = 100 - (Raster\ sand\ map + Raster\ silt\ map) \quad (2)$$

The purpose of applying this formulation was to preserve the compositional constraint of soil particle fractions, whereby sand, silt, and clay collectively sum to 100% at each pixel. Clay was therefore not interpolated independently. Modeling all three particle fractions separately using kriging could potentially lead to inconsistencies, including violations of the mass balance condition or unrealistic combinations at pixel level. By deriving clay as the residual fraction from the interpolated sand and silt maps, internal consistency among soil components was ensured while avoiding redundancy in spatial modeling.

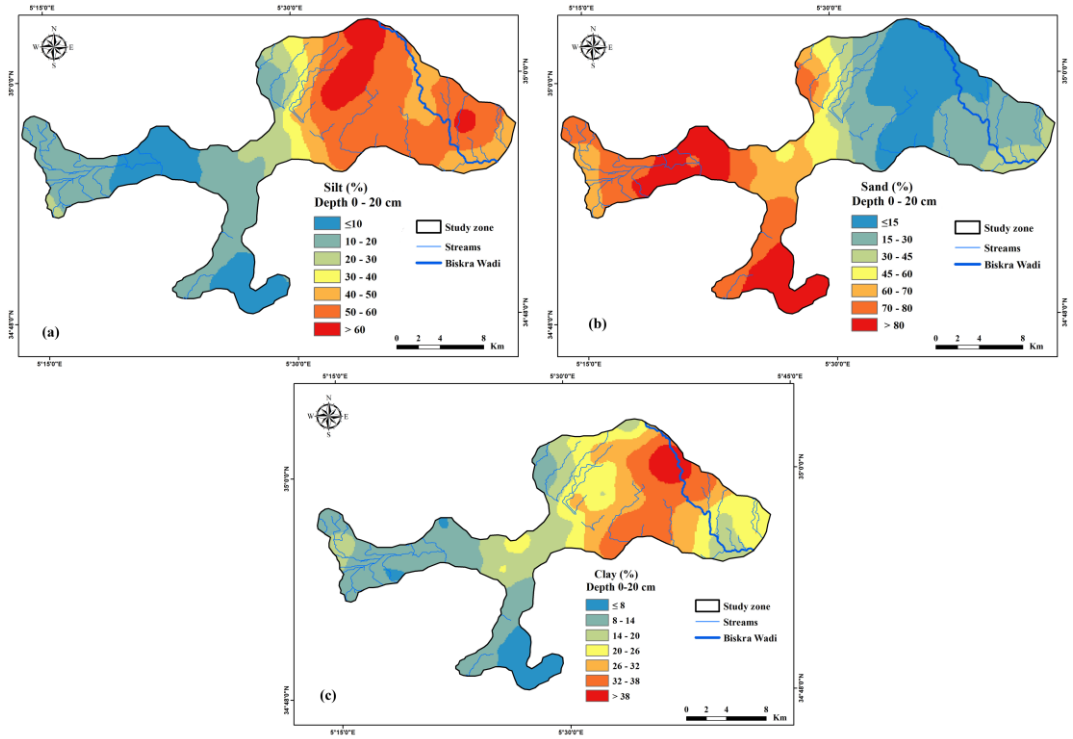

Figure 4. Spatial distribution maps of the soil particle fractions: (a) silt. (b) sand and (d) clay

2.4. Cross-validation and prediction accuracy

The accuracy of the interpolation results is assessed using the LOOCV method, employing the following indices:

$$ME = \frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)] \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[Z^*(x_i) - Z(x_i)]^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Z(x_i) - Z^*(x_i))}{\sum_{i=1}^{n}(\overline{Z(x)} - Z^*(x_i))} \quad (5)$$

Where:

- *ME* represents the mean error, which ideally approaches 0.
- *RMSE* denotes the root mean square error, aiming for minimum values.
- R^2 is the coefficient of determination, ranging from 0 to 1.
- *n* is the number of samples.
- $Z^*(x_i)$ signifies the estimated variable value.
- $Z(x_i)$ represents the measured variable value.
- $\sigma(x_i)$ indicates the estimation variance.

The accuracy of the results of texture classification map obtained by DTC is assessed using d Cohen's coefficient (Kappa coefficient = K), which can explain the agreement between the calculated and predicted classes (Kundel and Polansky, 2003; Hateffard et al. 2019). The coefficient is calculated using SPSS software V. 22 and the equation is defined as:

$$Kappa\ coefficient\ (K) = \frac{P_0 + P_e}{1 - P_e} \quad (6)$$

where, P_0 is the probability of observed agreement between determined and predicted classes, and P_e is the probability of agreement when two classes are independent. The strength of agreement depends on the K value, with classifications as follows: poor agreement (K<0), slight agreement (0 − 0.20), fair agreement (0.21 − 0.40), moderate agreement (0.41 − 0.60), substantial agreement (0.61 − 0.80), and almost perfect (0.81 − 1) (Landis and Koch 1977).

2.5. Implementation of decision tree classifiers (DTC)

Based on the United States Department of Agriculture (USDA) textural soil classifications (Table 1) (USDA 1987), four decision tree classifiers (DTCs) (Figure 5) are implemented to map soil texture using spatial distribution maps of silt, clay, and sand. The main idea is to transition from point-by-point determination of soil texture using the USDA texture triangle to generate a continuous soil texture map. Due to the intricate conditions and interrelations among soil particle fractions, generation of a single DTC is impractical to encompass all texture classes.

The first DTC (Figure 5a) meets the conditions of texture 1 (sand), texture 2 (Loamy sand), and texture 3 (Sandy loam). The second DTC (Figure 5b) satisfies the conditions of texture 7 (Sandy clay loam), texture 10 (Sandy clay), texture 11 (Silty clay), and texture 12 (Clay). The third DTC (Figure 5c) covers the conditions of texture 6 (Silt), texture 8 (Clay loam), texture 9 (Silty clay loam), the second condition of texture 3, and both conditions of texture 5 (Silt loam). The fourth DTC (Figure 5d) meets only the condition of texture 4 (Loam).

These four DTCs are executed in ArcGIS 10.8 based on pixel values of the three raster maps of soil particle fractions, using the Con function as follows:

$$\begin{aligned} DTC\ "A" = Con(&"Silt" + 1.5 \\ &* "Clay" \\ &< 15{,}1, Con("Silt" \\ &+ 2 * "Clay" \\ &< 30{,}2, Con("Sand" \\ &> 52, Con("Clay" \\ &< 20, Con("Clay" \\ &\geq 7{,}3{,}0){,}0){,}0))) \end{aligned} \quad (7)$$

$$\begin{aligned} DTC\ "B" = Con(&"Sand" \\ &> 45, Con("Clay" \\ &\geq 35{,}10, Con("Clay" \\ &\geq 20, Con("Silt" \\ &< 28{,}7{,}0){,}0)), Con("Clay" \\ &\geq 40, Con("Silt" \geq 40{,}11{,}12){,}0)) \end{aligned} \quad (8)$$

$$\begin{aligned} DTC\ "C" = Con(&"Clay" \\ &\geq 12, Con("Clay" < 27, Con("Silt" \\ &\geq 50{,}5{,}0), Con("Clay" \\ &< 40, Con("Sand" \\ &\leq 20{,}9, Con("Sand" \leq 45{,}8{,}0)){,}0) \end{aligned} \quad (9)$$

$$\begin{aligned} DTC\ "D" = Con(&"Sand" \\ &\leq 52, Con("Silt" \\ &\geq 28, Con("Silt" \\ &< 50, Con("Clay" \\ &< 27, Con("Clay" \\ &\geq 7{,}4{,}0){,}0){,}0){,}0){,}0) \end{aligned} \quad (10)$$

Here, DTC "A", DTC "B", DTC "C", and DTC "D" correspond to the four developed decision trees as depicted in Figure 5. "Silt", "Clay", and "Sand" represent the raster maps of the silt, clay, and sand soil fractions, respectively.

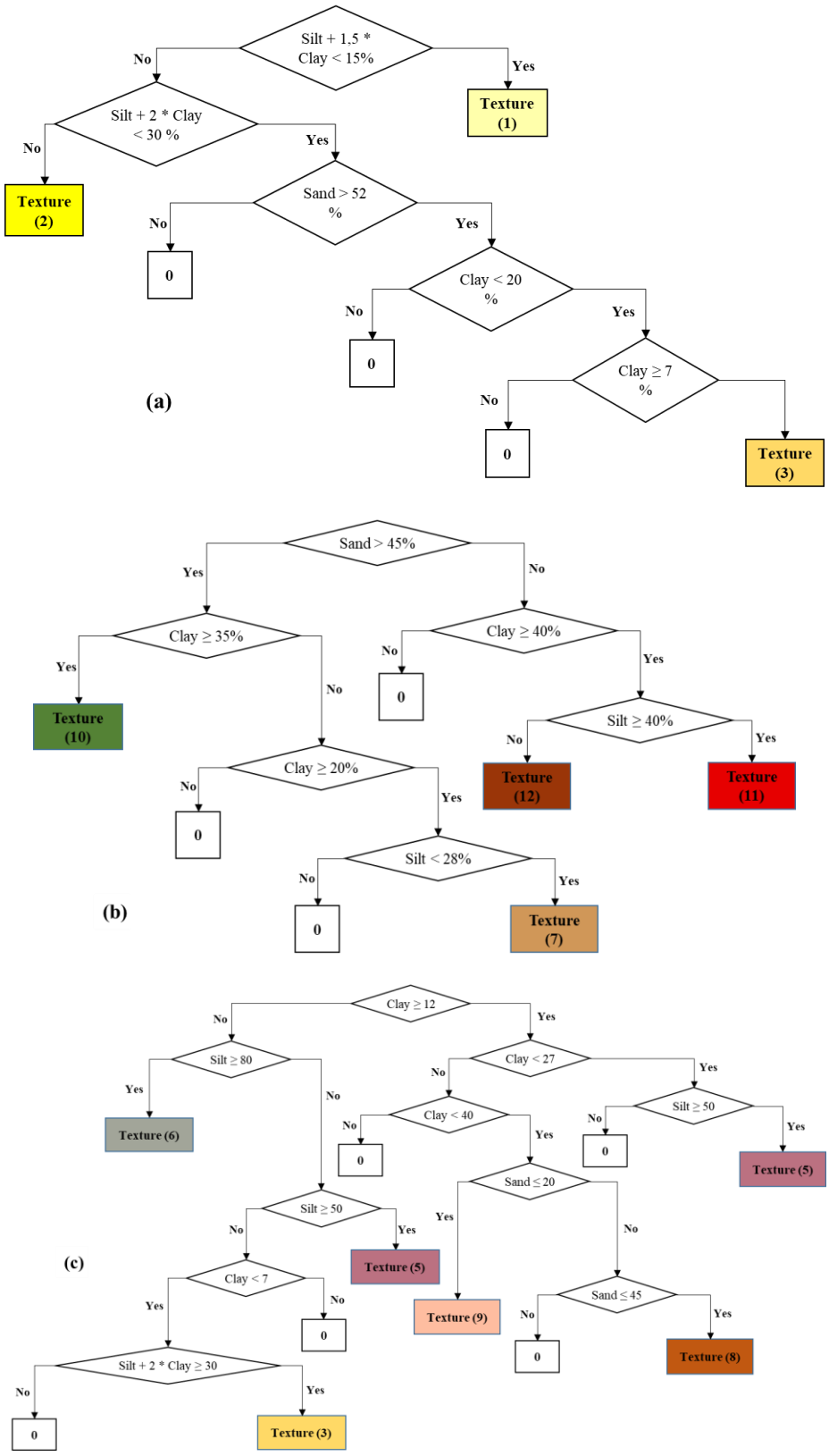

Figure 5. Decision Tree Classifiers [(a) DTC "A", (b) DTC "B", (c) DTC "C", (d) DTC "D"]

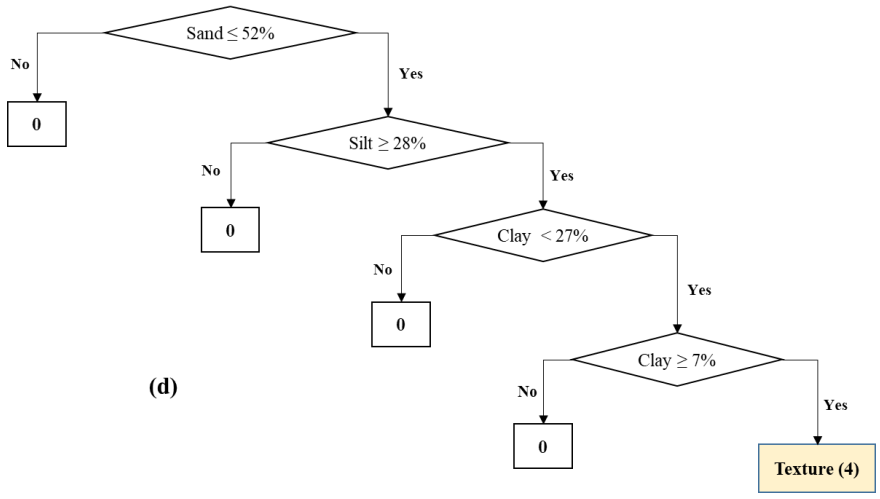

Figure 5. (Continued)

Execution of these DTCs generates four partial texture maps as in Figure 6, where a value from 1 to 12 is assigned to each pixel of the generated map when a condition of textural soil classification is satisfied, while the value 0 indicates that the condition does not fit any classification. The final texture map (Figure 7) is obtained by summing the four partial raster maps. The absence of the value 0 in the final map indicates that all conditions of the USDA textural soil classification are considered in the four DTC's.

$$Final\ texture\ map = DTC\ "A" + DTC\ "B" + DTC\ "C" + DTC\ "D" \quad (11)$$

Table 1. USDA soil texture classes and their assigned numbers and conditions

Soil texture	Numerical Number	Condition (USDA, 1987)
Sand	1	(silt + 1.5 × clay) < 15%
Loam sand	2	(silt + 1.5 × clay) ≥ 15% & (silt + 2 × clay) < 30%
Sandy loam	3	(clay ≥ 7% & clay < 20% & sand > 52% & (silt + 2 × clay) ≥ 30%) or (clay < 7% & silt < 50% & (silt + 2 × clay) ≥ 30%)
Loam	4	clay ≥ 7% & clay < 27% & silt ≥ 28% & silt < 50% & sand ≤ 52%
Silt loam	5	(silt ≥ 50% & clay ≥ 12% & clay < 27%) or (silt ≥ 50% & silt < 80% & clay < 12%)
Silt	6	silt ≥ 80 & clay < 12%
Sandy clay loam	7	clay ≥ 20% & clay < 35% & silt < 28% & sand > 45%
Clay loam	8	clay ≥ 27% & clay < 40% & sand > 20% & sand ≤ 45%
Silty clay loam	9	clay ≥ 27% & clay < 40% & sand ≥20%
Sandy clay	10	clay ≥ 35% & sand > 45%
Silty clay	11	clay ≥ 40% & silt ≥ 40%
Clay	12	clay ≥ 40% & sand ≥ 45% & silt < 40%

3. Results and discussion

3.1. Descriptive statistics

Descriptive statistics presented in Table 2 provide an overview of the soil particle-size fractions (%) analyzed in this study. Considerable variability in mineral fractions (clay, silt, and sand) was observed across the study area. Clay content ranged from 0% to 47.5%, silt from 2.23% to 89.79%, and sand from 1.55% to 92.5%. The mean values were 19.19%, 31.46%, and 49.34% for clay, silt, and sand, respectively. The relatively high coefficients of variation (CV), 66.72% for clay, 74.80% for silt, and 64.45% for sand, indicate substantial spatial heterogeneity of soil particle fractions within the study area.

Skewness and kurtosis values suggest moderate deviations from normality, particularly for silt and clay, while sand exhibited near-symmetrical distribution but a flatter-than-normal shape. The Shapiro-Wilk test confirmed statistically significant deviations from normality ($p < 0.05$) for all fractions. However, these deviations were not extreme and were considered acceptable for subsequent geostatistical analysis.

3.2. **Variography and Cross-validation** Figure 3 presents the best-fitted theoretical semivariograms and their associated parameters for the selected interpolation method (OK). Although several interpolation techniques (OK, SK, and IDW) were evaluated, only the semi variograms corresponding to the best-performing method (OK) are shown for clarity. The experimental semi variograms for sand and silt were best fitted using spherical models.

Table 2. Descriptive statistics and normality assessment of soil particle fractions

Variable	N	Min	Max	Mean	SD	CV (%)	Skewness	Kurtosis	W	p-value
Clay (%)	68	0	47.5	19.19	12.80	66.72	0.445	−1.017	0.932	0.001
Silt (%)	68	2.23	89.79	31.46	23.53	74.80	0.627	−0.553	0.899	< 0.001
Sand (%)	68	1.55	92.5	49.34	31.80	64.45	−0.134	−1.628	0.870	< 0.001

N: Number of samples. Min: Minimum. Max: Maximum. SD: Standard deviation. CV: Coefficient of variation. Shapiro-Wilk test: W

For sand, the nugget effect (C_0) and partial sill (C) were 106.833 and 774.921, respectively, with a range of 33366.100 m. For silt, the C_0 and C values were 157.889 and 1409.687, respectively, with a range of 33311.200 m. The nugget-to-sill ratios ($C_0/(C_0 + C)$) were 12.121% for sand and 10.070% for silt, indicating strong spatial dependence according to Cambardella et al. (1994). These results confirm that spatial continuity of the particle fractions is well defined within the study area.

LOOCV results are summarized in Table 3. The ME values were close to zero, indicating unbiased predictions. The RMSE values were 15.930 for sand and 13.110 for silt, while the R^2 values were 0.758 and 0.687, respectively. These statistics demonstrate acceptable predictive performance and support the generation of continuous spatial distribution maps within the proposed framework

Table 3. Variogram model parameters and LOOCV statistics for interpolated soil particle fractions

			Variography					LOOCV		
Depth (cm)	Soil fraction	Method	Model	Nugget (C_0)	Partial sill (C)	Range (m)	Nugget/Sill (%)	ME	RMSE	R^2
0-20	Sand	OK	Spherical	106.833	774.921	33366.100	12.121	**0.096**	**15.930**	**0.758**
		SK	Spherical	155.489	1233.751	28167.500	11.192	0.722	15.970	0.745
		IDW	/	/	/	/	/	0.054	16.676	0.722
	Silt	OK	Spherical	157.889	1409.687	33311.200	10.070	**0.095**	**13.110**	**0.687**
		SK	Spherical	106.720	720.448	30616.630	12.903	0.002	13.239	0.679
		IDW	/	/	/	/	/	-0.32	13.246	0.679

3.3. Spatial distribution of soil particles (silt, clay and sand)

Figure 4 illustrates the spatial distribution of soil particle fractions (silt, sand, and clay) at a depth of 0-20 cm. Figures 4a, 4b, and 4c are utilized to implement the DTCs on the GIS platform. The silt and clay prediction maps depict medium to higher values in the east and northeast regions of the study area (30% to 89.79% silt, 14% to 47.5% clay), and lower values in the west and southwest regions (less than 30% silt, less than 14% clay). Conversely, the sand prediction map reveals higher content in the west and southwest areas (over 60% sand), and medium to lower content in the central and eastern regions. These three maps are important practical requirements as inputs for the various decision tree classifiers (DTCs) implemented within ArcGIS using Eqs. 7, 8, 9, and 10.

While the generated maps reveal clear spatial trends, it is important to acknowledge that the number of soil samples (n = 68) may be considered moderate relative to the spatial extent of the study area. Geostatistical analysis revealed strong spatial dependence for both sand and silt fractions, as indicated by low nugget to sill ratios (10–13%). In geostatistical applications, interpolation reliability depends not only on sample size but also on spatial structure and sampling distribution. The strong spatial continuity observed in this study supports the adequacy of the sampling configuration for regional scale mapping. Nevertheless, lower sampling density may limit the detection of fine-scale variability and increase uncertainty near texture-class boundaries. Increasing sample density could potentially improve local prediction accuracy and reduce classification uncertainty, particularly in transitional zones.

3.4. Partial texture maps generated from DTC's and final texture map

Running the four DTCs above generates four partial texture maps, which are illustrated in Figure 6. The first map (Figure 6a) is generated by DTC "A" and illustrates the spatial distribution of three soil textures in the eastern and southeastern regions of the study area, namely, sand (Texture 1) covering 2.74% of the area, loamy sand (Texture 2) 11.76%, and sandy loamy (Texture 3) 29.33%. The second map (Figure 6b) produced by DTC "B" indicates the spatial distribution of two textures indicating sandy clay loam (Texture 7) in the central part of the study area, covering 1.10% of the area, and silty clay (Texture 11) in the eastern part, covering 1.20%. The third map (Figure 6c) obtained using DTC "C" displays three textures in the eastern region as silt loam (Texture 5) covering 13.17%, clay loam (Texture 8) 3.06%, and silty clay loam (Texture 9) 24.74%. The fourth map (Figure 6d) derived from DTC "D" depicts the spatial distribution of loam texture (Texture 4) in the central area, covering 12.89% of the total area.

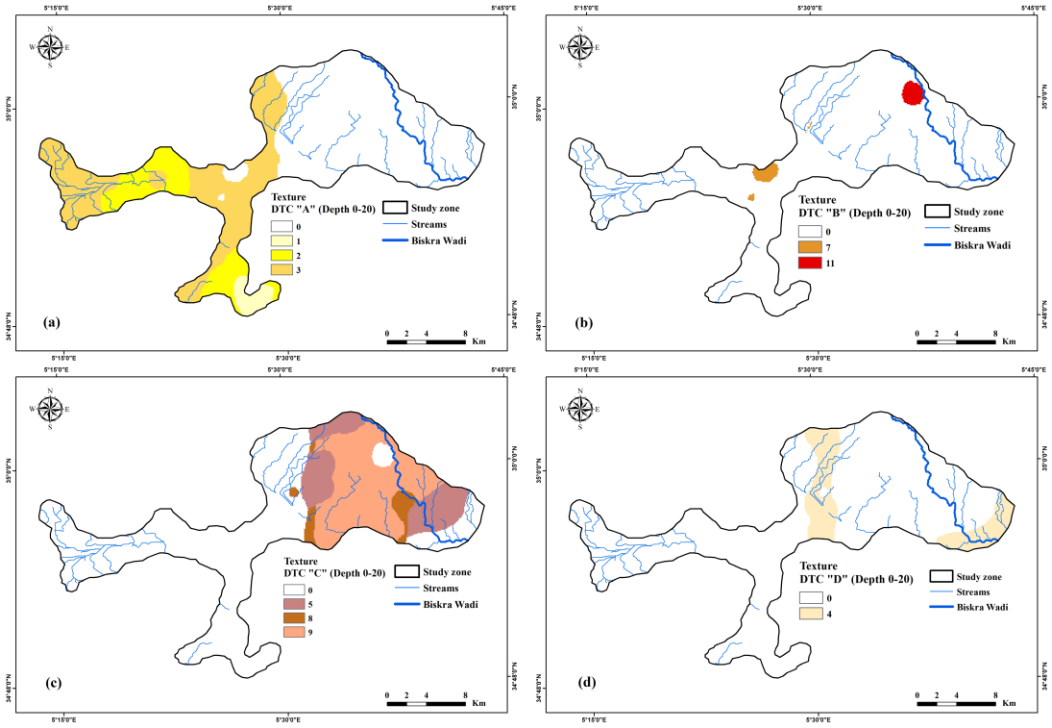

Figure 6. Partial texture maps generated using decision tree classifiers: (a) using DTC "A", (b) using DTC "B", (c)

The final map, derived from the summation of the four partial maps, is depicted in Figure 7. Soil texture classes are delineated as light (sand, loamy sand, and sandy loam) in the west and southwest, medium (loam and sandy clay loam) in the middle, and heavy (silt loam, clay loam, silty clay loam, and silty clay) in the east of the study area. The evaluation of soil texture classification using the proposed methodology indicates highly accurate results with an almost. perfect agreement (K = 0.898) between the determined and predicted classes. Although minor discrepancies were observed, a detailed examination suggests that these differences primarily occur near the boundaries of the USDA texture triangle. Because texture classes are defined by strict threshold values, small variations in sand, silt, or clay percentages may shift a pixel from one class to an adjacent class (e.g., sandy clay loam to sandy loam). Such shifts are particularly common in transitional zones where particle fractions lie close to class limits.

Furthermore, kriging interpolation inherently introduces spatial smoothing, which may attenuate local extremes and slightly modify particle fraction values at specific locations. Since clay content was derived from interpolated sand and silt maps to preserve the compositional constraint (sand + silt + clay = 100%), any interpolation uncertainty in sand and silt propagates into clay estimation, potentially amplifying small boundary shifts. Therefore, the observed discrepancies are mainly attributed to boundary sensitivity and interpolation-induced smoothing rather than systematic classification error. The deterministic decision tree classifiers strictly apply USDA rules at pixel level; thus, classification differences reflect minor numerical variations in interpolated fractions rather than inconsistencies in the classification procedure itself.

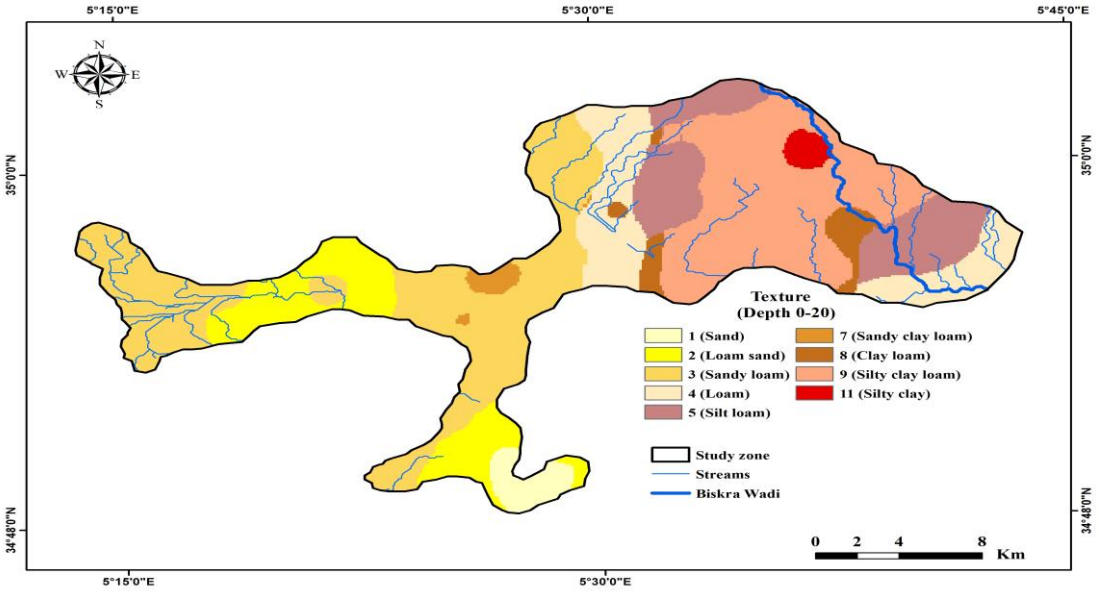

Figure 7. Final texture map of the study area

3.5. comparison with previous studies

Previous studies on soil texture mapping have utilized various ML models and environmental auxiliary variables to predict soil texture distribution. For instance, Song et al., (2023) employed different advanced ML algorithms and hyperspectral data for soil texture prediction, while Taghizadeh-Mehrjardi et al., (2020) used convolutional artificial neural networks (ANNs) and contextual data about the landscape for soil particle fractions mapping in Iran. Kaya et al., (2022) applied multinomial logistic regression and some environmental co-variables for digital mapping of soil texture in Turkey. These approaches generally aim to maximize predictive accuracy by integrating multiple explanatory variables derived from terrain attributes, climate, or remote sensing products.

However, in regions characterized by limited data availability, sparse sampling density, or restricted access to high-resolution auxiliary datasets, such approaches may encounter practical constraints related to data acquisition, preprocessing requirements, computational complexity, and model interpretability. Under these circumstances, simpler and more transparent frameworks may offer practical and operational advantages. In contrast, the present study deliberately relies exclusively on interpolated soil particle fractions followed by deterministic USDA-based classification rules. This approach does not aim to compete with multi-covariate ML models in terms of complexity, but rather proposes an alternative workflow emphasizing transparency, reproducibility, and methodological clarity. By separating the interpolation stage from the classification stage, the methodology ensures strict consistency with USDA texture criteria at the pixel level while preserving the compositional relationships among sand, silt, and clay fractions.

The proposed methodology is structured around two key principles:

- Combining accurate interpolation and decision trees: Utilizing OK ensures the generation of soil particle fractions maps, while decision tree classifiers effectively transform these maps into comprehensive soil texture maps.
- Efficiency and Practicality: The method requires only a limited number of soil sampling points and can be executed within a well-known GIS environment, making it accessible and practical for various applications.

The results demonstrate that, under conditions of moderate sampling density and well-defined spatial dependence, reliable soil texture maps can be generated without incorporating auxiliary environmental variables. Nevertheless, it is acknowledged that integrating additional covariates may further enhance local-scale predictions, particularly in highly heterogeneous landscapes. Therefore, the proposed methodology should be considered a practical and complementary alternative within the broader spectrum of digital soil mapping approaches.

4. Conclusion

This study presents a structured and reproducible methodology for mapping soil texture by combining geostatistical interpolation with fully deterministic decision tree classifiers based on USDA classification rules. By interpolating soil particle fractions to generate continuous spatial distribution maps of silt, clay, and sand, and subsequently applying decision tree classifiers, a spatially consistent and rule-based soil texture classification is achieved. The validation metrics (RMSE, R^2, and ME) demonstrate the robustness of the interpolation technique, while the high agreement between predicted and observed texture classes (K = 0.898) confirms the reliability of the overall framework. The findings suggest that this methodology is effective in producing reliable soil texture maps with practical applications in earth sciences and environmental management. By requiring only, a limited number of soils sampling points and leveraging GIS software, this approach offers a practical and efficient solution for soil texture mapping, particularly under conditions of moderate sampling density and clearly defined spatial dependence. The framework demonstrates that reliable soil texture maps can be generated without reliance on auxiliary environmental covariates or complex machine learning architectures, making it especially suitable for data-limited regions.

The implications of these results are significant for improving soil management practices, optimizing crop production, and enhancing environmental sustainability. This methodology provides a valuable tool for researchers and practitioners, facilitating better decision-making and contributing to advancements in soil science. Although increasing sampling density or incorporating additional environmental variables may further refine local-scale predictions, the proposed workflow represents a transparent and operationally efficient alternative within the broader context of digital soil mapping. For future research, development of an integrated decision-tree structure that consolidates the four classifiers into a unified framework could further streamline the methodology. This is not possible in the case in this paper due to the interaction between the conditions of the USDA soil texture triangle, but addressing this challenge could enhance the efficiency and applicability of the approach.

Author's contribution - Boudibi Samir: study conception and design, manuscript writing, data collection, modeling, mapping, and corresponding author; **Zineeddine Benguega and Fadlaoui Haroun**: Data collection and fieldwork; **Tarik Otman and Bouzidi Narimen**: Laboratory analyses support; **Sakaa Bachir and Aissaoui Azeddine:** Manuscript revision assistance; **Khomri Zinne-eddine**: Fieldwork support.

Data Availability Statement - The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests - The authors declare that they have no competing interests.

References

Abzalov, M. (2016). Variography. In: Applied Mining Geology. Modern Approaches in Solid Earth Sciences, vol 12. Springer, Cham. doi: 10.1007/978-3-319-39264-6_18

Ahmadi, S.H., & Sedghamiz, A. (2007). Geostatistical Analysis of Spatial and Temporal Variations of Groundwater Level. Environmental Monitoring and Assessment, 129, 277–294. doi: 10.1007/s10661-006-9361-z

Anthony, T., Nkwunonwo, U., Emmanuel, A., & Ganiyu O. (2025). Environmental and geostatistical modelling of soil properties toward precision agriculture. Discover Soil, 2, 59. doi: 10.1007/s44378-025-00083-y

Armstrong, M., & Boufassa, A. (1988). Comparing the robustness of ordinary kriging and lognormal kriging: Outlier resistance. Mathematical Geology, 20, 447–457. doi: 10.1007/BF00892988

Arslan, H. (2012). Spatial and temporal mapping of groundwater salinity using ordinary kriging and indicator kriging: The case of Bafra Plain, Turkey. Agricultural Water Management, 113, 57–63. doi: 10.1016/j.agwat.2012.06.015

Bajjali, W. (2023). ArcGIS Pro and ArcGIS Online Applications in Water and Environmental Sciences. Springer Nature, Switzerland. doi : 10.1007/978-3-031-42227-0

Benchetrit, M. (1956). Les sols d'Algérie. Revue de Géographie Alpine, 44, 749–761. doi : 10.3406/rga.1956.1790

Bernoux, M., Arrouays, D., Cerri, C.E.P., & Cerri, C.C. (2007). Regional organic carbon storage maps of the western Brazilian Amazon based on prior soil maps and geostatistical interpolation. In: Developments in Soil Science. Elsevier B.V., pp 497–507. doi: 10.1016/S0166-2481(06)31037-9

Bidkhani, G.O.N., & Mobasheri, M.R. (2018). Influence of soil texture on the estimation of bare soil moisture content using MODIS images. European Journal of Remote Sensing, 51, 911–920. doi: 10.1080/22797254.2018.1514986

Boudibi, S. (2021). Modeling the Impact of Irrigation Water Quality on Soil salinieation in an Arid Region, Case of Biskra. 176p. doi : 10.13140/RG.2.2.12406.93768

Bradaï, A., Douaoui, A., Bettahar, N., & Yahiaoui, I. (2016). Improving the Prediction Accuracy of Groundwater Salinity Mapping Using Indicator Kriging Method. Journal of Irrigation and Drainage Engineering, 142, 11. doi: 10.1061/(ASCE)IR.1943-4774.0001019

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A. E. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. Soil Science Society of America Journal, 58, 1501–1511. doi: 10.2136/sssaj1994.03615995005800050033x

Chau, J.F., Bagtzoglou, A.C., & Willig, M.R. (2011). The Effect of Soil Texture on Richness and Diversity of Bacterial Communities. Environment Forensics, 12, 333–341. doi: 10.1080/15275922.2011.622348

De Caires, S. A., Martin, C. S., Atwell, M. A., Kaya, F., Wuddivira, G. A., & Wuddivira, M. N. (2025). Advancing soil mapping and management using geostatistics and integrated machine learning and remote sensing techniques: a synoptic review. Discover Soil, 2(53), 1–29. doi: doi: 10.1007/s44378-025-00082-z

Durand, M.J.H., & Barbut, M.M. (1938). Carte de reconnaissance des sols d'Algérie : Biskra. Service Geographique de l'Armée.

ESRI., (2024). Using cross validation to assess interpolation results.

Grunwald, S. (2009). Geoderma multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma, 152, 195–207. doi: 10.1016/j.geoderma.2009.06.003

Han, J., Kamber, M., Pei, J. (2012). Data Mining: Concepts and Techniques, 3rd edn. Elsevier Inc, Waltham. doi: 10.1016/B978-0-12-381479-1.00022-8

Hateffard, F., Dolati, P., Heidari, A., & Zolfaghari, A.A. (2019). Assessing the performance of decision tree and neural network models in mapping soil properties. Journal of Mountain Science, 16, 1833-1847. doi: 10.1007/s11629-019-5409-8

Johnston, K., Ver Hoef, J.M., Krivoruchko, K., & Lucas, N. (2001). Using ArcGIS Geostatistical Analyst. ESRI, Redlands (California).

Kaya F., Başayiğit L., Keshavarzi A., & Francaviglia R. (2022). Digital mapping for soil texture class prediction in northwestern Türkiye by different machine learning algorithms. Geoderma Regional, 31, e00584. doi: 10.1016/j.geodrs. 2022.e00584

Kundel, H.L., & Polansky, M. (2003). Measurement of Observer Agreement. Radiology, 228,303–308. doi: 10.1148/radiol.2282011860

Lagacherie, P., McBratney, A.B., & Voltz, M. (2007). Digital Soil Mapping: An Introductory Perspective, 1st edn. Elsevier B.V., Kidlington

Landis, J.R., & Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33, 159–174. doi: 10.2307/2529310

Liao, K., Xu, S., Wu, J., & Zhu, Q. (2013). Spatial estimation of surface soil texture using remote sensing data. Soil Science and Plant Nutrition, 59, 488–500. doi: 10.1080/00380768.2013.802643

Maroufpoor, S., Fakhri-Fard, A., & Shiri, J. (2017). Study of the spatial distribution of groundwater quality using soft computing and geostatistical models. ISH Journal of Hydraulic Engineering, 25, 232–238. doi: 10.1080/09715010.2017.1474389

Minasny, B., & Mcbratney, A.B. (2015). Geoderma Digital soil mapping: A brief history and some lessons. Geoderma, 264, 301–311. doi: 10.1016/j.geoderma.2015.07.017

Oliver, M.A., & Webster, R. (2014). A tutorial guide to geostatistics : Computing and

modelling variograms and kriging. Catena 113, 56–69. doi: 10.1016/j.catena.2013.09.006

Pannatier, Y. (1996). VARIOWIN: Software for Spatial Data Analysis in 2D. Springer, New York. doi: 10.1007/978-1-4612-2392-4

Qu, L., Lu, H., Tian, Z., Schoorl, J.M., Huang, B., Liang, Y., Qiu, D., & Liang, Y. (2024). Spatial prediction of soil sand content at various sampling density based on geostatistical and machine learning algorithms in plain areas. Catena, 234, 107572. doi: 10.1016/j.catena.2023.107572

Karp, F.H.S., Adamchuk, V., Dutilleul, P., & Melnitchouck, A. (2024). Comparative study of interpolation methods for low-density sampling. Precision Agriculture. doi: 10.1007/s11119-024-10141-0

Rahmati, O., Falah, F., Naghibi, S. A., Biggs, T., Soltani, M., Deo, R. C., Cerdà, A., Mohammadi, F., & Bui, D. T. (2019). Land subsidence modelling using tree-based machine learning algorithms. Science of the total environment, 672, 239-252. doi: 10.1016/j.scitotenv.2019.03.496

Ravikumar, V. (2022) Sprinkler and Drip Irrigation: Theory and Practice. Springer. Singapore. doi: 10.1007/978-981-19-2775-1

Santra, P., Kumar, M., & Panwar, N. (2017). Digital soil mapping of sand content in arid western India through geostatistical approaches. Geoderma Regional, 9,56–72. doi: 10.1016/j.geodrs.2017.03.003

Seyedmohammadi, J., Navidi, M.N., & Esmaeelnejad, L. (2019). Geospatial modeling of surface soil texture of agricultural land using fuzzy logic, geostatistics and GIS techniques. Communications in Soil Science and Plant Analysis, 50, 1452–1464. doi: 10.1080/00103624.2019.1626870

Song, Q., Gao, X., Song, Y., Li, Q., Chen, Z., Li, R., Zhang, H., & Cai, S. (2023). Estimation and mapping of soil texture content based on unmanned aerial vehicle hyperspectral imaging. Scientific Reports, 13, 1–17. doi: 10.1038/s41598-023-40384-2

Şen, Z. (2020). Earth Systems Data Processing and Visualization Using MATLAB. Springer Nature, Cham. doi: 10.1007/978-3-030-01542-8

Szczerbicki, E. (2001). Management of Complexity and Information Flow. In: Gunasekaran A (ed) agile manufacturing: the 21 st century competitive strategy. Elsevier Ltd, Kidlington, p 810. doi: 10.1016/B978-008043567-1/50013-9Taghizadeh-

Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafilis, J., & Omid, M. (2014). Digital Mapping of Soil Classes Using Decision Tree and Auxiliary Data in the Ardakan Region, Iran. Arid Land Research and Management, 28, 37–41. doi: 10.1080/15324982.2013.828801

Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., & Schmidt, K. (2020). Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. Geoderma, 376, 114552. doi: 10.1016/j.geoderma.2020.114552

USDA. (1987). Soil Mechanics: USDA Soil Textural Classification Study Guide. USDA Soil Conservation Service, Washington DC.

Wadoux, A.M.J-C., Minansy, B., & McBratney, A.B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews, 210, 103359. doi: 10.1016/j.earscirev.2020.103359

Wani, O. A., Sharma, V., Kumar, S. S., Malik, A. R., Pandey, A., Devi, K., Kumar, V., Gairola, A., Yadav, D., Valente, D., Petrosillo, I., & Babu, S. (2024). Geostatistical modelling of soil properties towards long-term ecological sustainability of agroecosystems. Ecological Indicators, 166, 112540. doi: 10.1016/j.ecolind.2024.112540

Webster, R., & Oliver, M.A. (2007). Geostatistics for Environmental Scientists. John Wiley & Sons, Southern Gate

Wu, G., Shen, D., & Sabuncu, M.R. (2016). Machine Learning and Medical Imaging. Academic Press. Amsterdam

Yousif, B.S., Mustafa, Y.T., & Fayyadh, M.A. (2023). Digital mapping of soil-texture classes in Batifa, Kurdistan Region of Iraq, using machine-learning models. Earth Science Informatics, 16, 1687–1700. doi: 10.1007/s12145-023-01005-8

Zhao, C.X., Jia, L.H., Wang, Y.F., Wang, M. L., McGiffen, Jr.M.E. (2015). Effects of Different Soil Texture on Peanut Growth and Development and Development. Communications in Soil Science and Plant Analysis, 46, 2249–2257. doi: 10.1080/00103624.2015.1059845.