Earth Science Informatics (2026) 19:28
https://doi.org/10.1007/s12145-026-02077-y

RESEARCH

Adaptive neuro-fuzzy inference system for predicting sandy soils bulk density

Dian Lourençoni[1] · Alessandra Monteiro Salviano[2] · Nelci Olszevski[1] · Janielle Souza Pereira[3] · Edmo Henrique Martins Cavalcante[1]

Received: 31 October 2025 / Accepted: 26 January 2026

Abstract
Soil compaction, often the result of intensive mechanization, is a major concern because it increases the bulk density of the soil, fundamentally altering its structure and reducing its capacity to support essential ecological and agricultural functions. Therefore, soil monitoring is essential, particularly through its main quality indicator, bulk density. However, the analysis of this attribute requires the collection of soil samples with subsequent laboratory analysis, which demands time and qualified labor. Thus, this study aimed to develop a neuro-fuzzy model to estimate the bulk density of sandy soils based on variations in the proportion of its particle size fractions. To achieve this, undisturbed samples were collected from soils in four municipalities in northern Bahia, located along the shores of Lake Sobradinho. The proportion of particle size fractions in the composition of each sample was used to develop models that integrate fuzzy logic and neural networks. They employed the Takagi-Sugeno inference method and hybrid learning algorithms for parameter adjustment and bulk density predictions. The performance of the models was evaluated using error statistics, with the model with the highest accuracy in density predictions being selected, achieving an R^2 of 71%. It is important to emphasize that the neuro-fuzzy model did not require data stratification and can be applied to estimate the sandy soil density of any soil type sampled.

Keywords Bulk density · Estimation of bulk density · Neuro-fuzzy model · Sandy soil

Introduction

With frequent and intensive mechanization across all stages of agricultural cultivation, resulting from the increasing need to enhance crop productivity and work efficiency, the machinery used in operations is becoming progressively heavier and more powerful (McPhee et al. 2020). In this scenario, soil compaction is a consequence of various production systems. It restricts root development (Bengough et al. 2006) due to higher soil density and penetration resistance, while also reducing air and water permeabilities. Moreover, changes in porosity and infiltration increase surface runoff, which, in severe cases, elevates the risk of water and wind erosion (Horn et al. 1995; Mileusnić et al. 2022; USDA 2023; Sturm et al. 2025).

In addition, soil compaction from farm machinery traffic significantly increases the bulk density and cone index (penetration resistance) by pressing mineral components closer together, thereby reducing the air and water volume (Raper, 2005). This compression reduces the total porosity and macropore volume, decreasing the air permeability, relative gas

Communicated by Hassan Babaie

✉ Edmo Henrique Martins Cavalcante
edmo.cavalcante@ufrpe.br

Dian Lourençoni
dian.lourenconi@univasf.edu.br

Alessandra Monteiro Salviano
alessandra.salviano@embrapa.br

Nelci Olszevski
nelci.olszevski@univasf.edu.br

Janielle Souza Pereira
janielle.engenheira@gmail.com

[1] Postgraduate Program in Agricultural Engineering, Federal University of São Francisco Valley (UNIVASF), Av. Antônio Carlos Magalhães, 510, Juazeiro, BA 48902-300, Brazil

[2] Brazilian Agricultural Research Corporation (Embrapa Semiárido), BR-428, Petrolina, PE 56302-970, Brazil

[3] Postgraduate Program in Agricultural Engineering, Federal University of Lavras (UFLA), Trevo Rotatório Professor Edmir Sá Santos, MG 37203-202 Lavras, Brazil

Published online: 26 February 2026

diffusivity, and hydraulic conductivity by 55–82% (Kuncoro et al., 2014).

Increased mechanical impedance also restricts root penetration and oxygen availability, often stopping root growth completely at soil strengths above 2.96 MPa. Diminished pore space reduces root length density and alters root distribution patterns, while the resulting poor infiltration creates surface runoff pathways that aggravate soil erosion (Raper, 2005). Indirectly, compaction exacerbates soil erosion by reducing infiltration rates and creating runoff pathways, with wheel ruts functioning as conduits for surface water flow (Raper, 2005; Jensen et al., 2025).

Therefore, monitoring soil quality through the evaluation of its physical attributes has assumed a crucial role in the sustainability of agricultural systems and the maintenance of productive capacity (Herrick 2000). According to Cunha et al. (2002) and Han and Wang (2023), soil bulk density (BD) is traditionally used as a soil quality indicator because of its relationship with other attributes, such as porosity and moisture.

Thus, when the structural organization of soil undergoes degradation, the immediate effect is an increased BD and, consequently, alteration of porosity (Barbosa et al. 2020) with increased mechanical impediment to root growth (Gomes Junior et al. 2022). In this regard, granulometry significantly influences the compressive behavior of different soils owing to the rearrangement of particles (Zheng et al. 2023; Chen et al. 2024). This characteristic translates into load-bearing capacity and soil resistance to deformation, which substantially affects machinery and implements traffic and crop development (Labelle and Kammermeier 2019).

When machinery loads exceed the precompression stress threshold, it results in compaction that increases the bulk density and reduces the porosity (Hamza and Anderson 2005). It substantially affects machinery traffic by elevating rolling resistance and fuel consumption (Rima et al. 2024; Jensen et al., 2025), while also restricting operational timing due to increased vulnerability at high soil moisture content (Hamza and Anderson 2005). For crop development, the resulting soil deformation creates mechanical impedance that severely restricts root elongation and diminishes water infiltration, nutrient availability, and aeration, ultimately reducing yields (Forster et al., 2020).

The relationship between bulk density and soil granulometry is complex, as particle size, shape, and arrangement significantly influence the degree of packing and the resulting pore structure (Robinson et al., 2022; Zhang et al., 2022). Soils with a higher content of coarse particles, such as sand, tend to exhibit naturally higher bulk densities, due to the relatively simple packing of these granular components (Horn et al. 1995). Conversely, soils rich in finer particles, specifically clay and organic matter, generally have lower bulk densities (Raper, 2005). Clay and silt particles promote aggregation, which creates greater total porosity and a higher volume of macropores and micropores, thus reducing the bulk density (Robinson et al., 2022).

The method commonly used to determine BD consists of collecting samples with an undisturbed structure using a volumetric ring of known volume (Teixeira et al. 2017). Although low-cost and accessible, this method may alter soil organization depending on the sampling tool, volume of soil collected, and operator skill (Matsinhe et al. 2019), masking actual soil conditions. In addition to the sample collection process, laboratory analysis requires time and qualified personnel.

Therefore, pedotransfer functions are commonly used to estimate BD from easily available and measurable soil attributes (Bernoux et al. 1998; Benites et al. 2007; Barros and Fearnside 2015; Souza et al. 2016; Beutler et al. 2017). However, their predictive potential may be limited to experimental conditions or require data stratification by horizon or soil class (Pádua et al. 2015).

Owing to the complexity of the interactions involved, decision support systems can be used to assist actions (Omomule et al. 2020). In this context, mathematical and computational modelling has been successfully used to describe BD behavior (Bernoux et al. 1998; Benites et al. 2007; Barros and Fearnside 2015; Pádua et al. 2015; Souza et al. 2016; Beutler et al. 2017).

However, classical mathematical methods may not produce satisfactory results because of the subjectivity of information and the presence of environmental factors in systems. Alternatively, the application of methodologies based on artificial intelligence, such as fuzzy logic and neuro-fuzzy, is especially useful in dealing with complex problems and imprecise information and in easily translating them into mathematical languages (Halder et al. 2022). These systems allow consistent modelling of cause-and-effect relationships from available data and specialist knowledge. Furthermore, some rules explicitly explain the dependency and relationships between the model factors, assisting in decision-making (Novák 2006).

Thus, various studies have achieved good performance utilizing fuzzy modelling applied to soil response simulations, such as soil attribute mapping (Zhu et al. 2010; Beucher et al. 2014), landslide susceptibility analysis (Akgun et al. 2012; Osna et al. 2014), quality analysis of soils undergoing recovery (Branco et al. 2021), and its applicability in agriculture (Godinho et al. 2022). However, studies evaluating BD prediction using intelligent models such as fuzzy logic and neuro-fuzzy are lacking. Thus, this study aimed to develop a neuro-fuzzy model to estimate the bulk density of sandy soils based on variations in the proportion of its particle size fractions and their complex relationships.

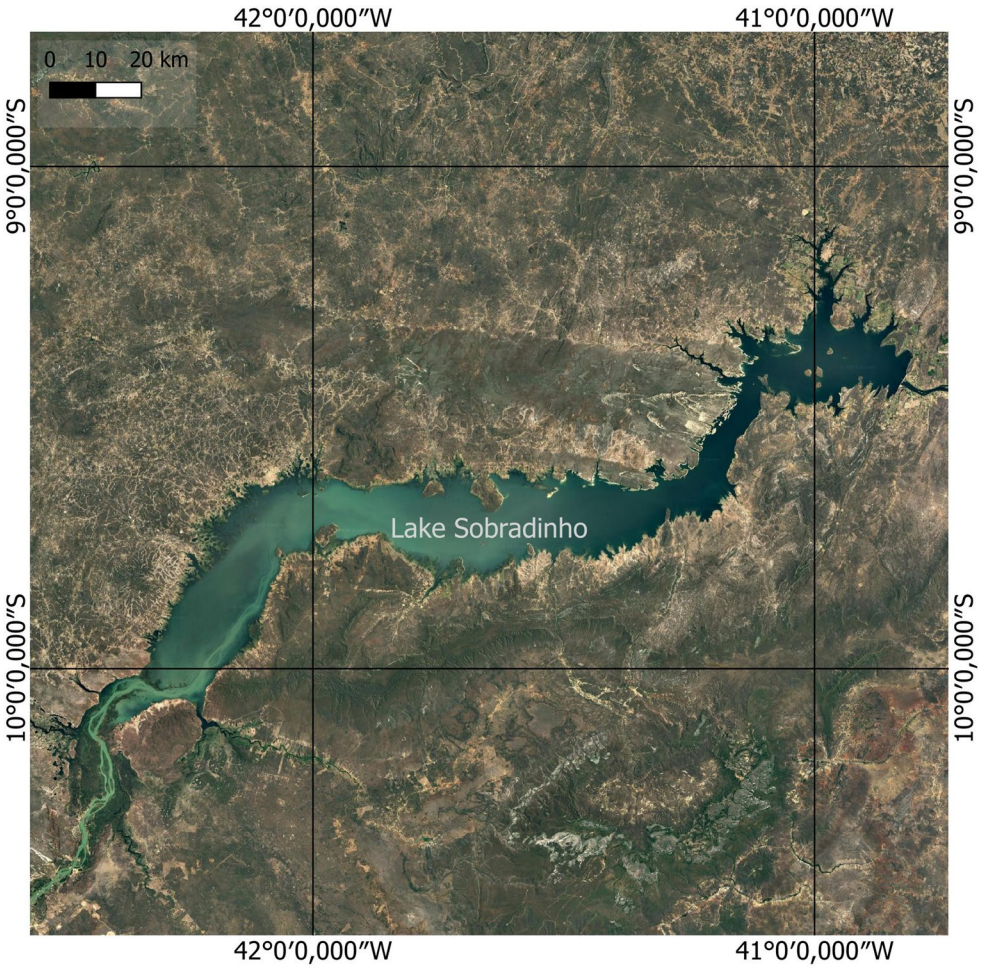

Fig. 1 Region from which samples were collected

Materials and methods

Sites and soil sampling

This study was conducted in two stages: an experimental stage involving soil sample collection in four municipalities in northern Bahia and a neuro-fuzzy model training and, validation, and test stages.

Sampling was conducted in the Lower Middle São Francisco region, covering the municipalities of Sobradinho, Casa Nova, Remanso, and Sento Sé. These areas are located along the shores of Lake Sobradinho in the northern part of Bahia State (Fig. 1).

The region has a hot, semi-arid climate (BSh type, according to the Köppen classification), with annual rainfall between 500 mm and 900 mm. Precipitations are concentrated from December to March, followed by prolonged drought periods. The natural vegetation hyperxerophilic caatinga (seasonally tropical dry forest) (Queiroz, et al., 2017), typical of the steppe-savanna biome. The study sites were chosen because they represent the soil types most commonly used for agriculture in the region (Table 1), such as Acrisols, Cambisols, Planosols, Arenosols, Ferralsols and Luvisols (Arcoverde et al. 2015). Soils used primarily for agricultural purposes, mainly for the cultivation of onions, bananas, watermelons, melons, and cassava.

Table 1 Soil classification and sampling depth

Municipality	Soil classification (WRB)	Sampling depth (cm)
Sobradinho	Cambisols, Acrisols	0–150
Casa Nova	Acrisols, Ferralsols, Arenosols, Planosols	0–210
Remanso	Ferralsols, Arenosols	0–200
Sento Sé	Acrisols, Arenosols, Planosols	0–200

Laboratory analysis

Bulk density (BD)

The BD of each sample was analyzed using the volumetric ring method. This method involves collecting an undisturbed soil sample using a metal ring (or cylindrical core sampler) of a precisely known internal volume (Rodriguez et al. 2025; Teixeira et al. 2017). Rings of 0.05 m height and

0.0508 m (2 in) diameter were utilized, totaling a volume of 101.33 cm^3. The undisturbed samples were oven-dried at 105–110 °C for 24 h. BD was calculated using Eq. 1.

$$BD = m/V \tag{1}$$

where BD is the bulk density (kg dm^{-3}); m is the dry weight (g), and V is the internal volume of the Kopecky ring (cm^{-3}).

Particle size distribution

To determine the proportions of clay, silt, and sand fractions, particle size analysis was performed according to Teixeira et al. (2017), in addition to fractionation of sand subclasses according to the United States Department of Agriculture (USDA) classification: very fine sand (VFS), from 0.10 mm to 0.05 mm; fine sand (FS), from 0.25 mm to 0.10 mm; medium sand (MS), from 0.50 mm to 0.25 mm; coarse sand (CS), from 0.50 mm to 1.00 mm; and very coarse sand (VCS), from 2.00 mm to 1.00 mm in diameter.

The particle size fractions and bulk density of each sample were summarized using descriptive statistics, such as the mean, standard deviation (SD), and coefficient of variation (CV), and illustrated using box plots. In this graph, the box represents the interquartile range, showing the spread of the middle 50% of the data points, and its central line represents the median. The relationships among these variables were examined using the Pearson correlation coefficient (r) and its significance level.

Neuro-fuzzy modelling

Neuro-fuzzy systems were developed, and their parameters were adjusted based on the collected soil sample data. The response variable was soil density (BD), and the model input variables were the values of very coarse sand (VC), coarse sand (CO), medium sand (ME), fine sand (FI), very fine sand (VF), clay (CL), and silt (SI).

For neuro-fuzzy system development, these data were randomly distributed into two groups, with 69.8% (257 observations) for training (model parameter adjustments), 15.2% (56 observations) for validation, and 14.9% (55 observations) for testing. This proportion provided sufficiently large dataset for the system to learn the complex nonlinear relationships among the seven particle size inputs while still preserving a substantial, independent subset to assess predictive accuracy and avoid overfitting. This approach balances model complexity with data availability, providing reliable performance metrics without sacrificing the statistical power required for a high dimensional neuro-fuzzy model.

Models were developed in MATLAB® Fuzzy Toolbox platform, version 7.13.0.564 (R2011b) using the Takagi-Sugeno inference method (Takagi and Sugeno 1985) selected for its efficiency in previous studies (Lins et al. 2021; Santos et al. 2024). The method is based on IF-THEN rules, in which rule R_k is denoted by Eqs. 2 and 3, respectively.

$$\begin{aligned} &Rule R_1: \; lF\,(x_1 is\; A_1) \wedge (x_2 is B_2) \\ &THEN y_1 = \alpha_{10} + \alpha_{11}x_1 + \alpha_{12}x_2 \end{aligned} \tag{2}$$

$$\begin{aligned} &Rule R_2: \; lF\,(x_1 is\; A_2) \wedge (x_2 is B_2) \\ &THEN y_2 = \alpha_{20} + \alpha_{21}x_1 + \alpha_{22}x_2 \end{aligned} \tag{3}$$

Figure 2 illustrates the typical architecture of a neuro-fuzzy system with five layers for predicting unknown data, with inputs (x_1, …, x_7) representing the aforementioned particle size fractions and output f, soil density. In the figure, the circles represent fixed nodes, and the rectangles represent adaptive nodes, whose membership function and fuzzy rule parameters are optimized by the neural structure. Terms A, …, and G denote linguistic terms, each represented by 2

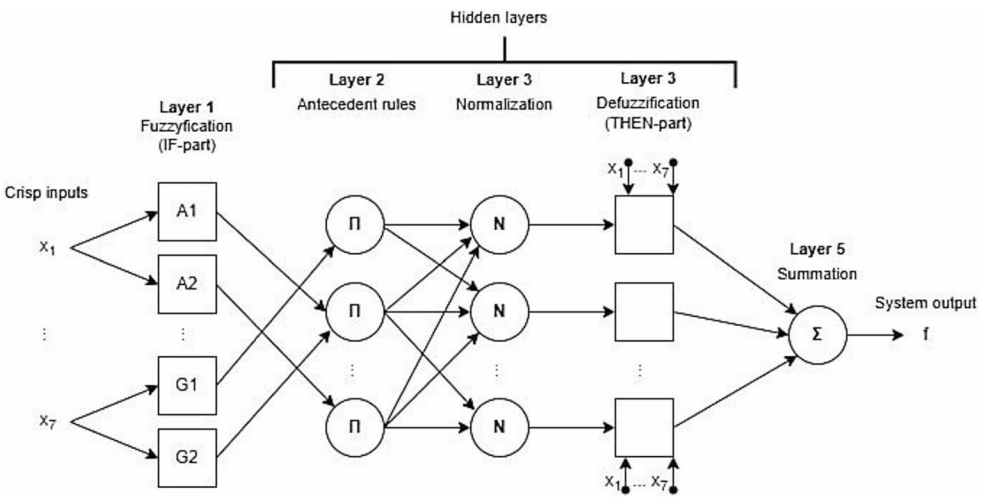

Fig. 2 ANFIS architecture. Source: Adapted from Jang (1993)

membership functions that translate, in mathematical terms, imprecise information from the real world (Mendel 2024).

Different membership functions (Gaussian, trapezoidal, and triangular), numbers of training sections, and optimization methods (backpropagation and hybrid) were also tested. The model with the lowest prediction error was selected. The generalization capacity was evaluated using an independent dataset that was not included in the training stage; the test set was used to assess the model's performance, allowing the calculation of error statistics in a consistent manner.

Model performance

Model performance, which is fundamental to increase confidence and enable model selection (Tedeschi 2006), was evaluated by comparing predicted and observed values from the test set (15% of the total dataset). Student's t-test was used to test the mean differences. Accuracy was quantified using bias, mean absolute error (MAE), coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), root mean squared logarithmic error (RMSLE), mean absolute percentage error (MAPE), and Nash–Sutcliffe efficiency coefficient (Nash and Sutcliffe 1970), calculated by Eq. 4 to 11, respectively.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i) \tag{4}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \tag{5}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N}(P_i - \overline{P}) \bullet (O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N}(P_i - \overline{P})^2 \bullet \sum_{i=1}^{N}(O_i - \overline{O})^2}}\right]^2 \tag{6}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2 \tag{7}$$

$$RMSE = \sqrt{MSE} \tag{8}$$

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[log(O_i + 1) - log(P_i + 1)]^2} \tag{9}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{O_i - P_i}{O_i}\right| \tag{10}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}\right] \tag{11}$$

where n is the total number of samples, P_i is the i-th predicted values, O_i is the i-th observed value, $\overline{P}$ is the average of predicted values, and $\overline{O}$ is the average of observed values.

Results and discussion

Physical characteristics of soils

The bulk density and particle size data used to develop the neuro-fuzzy models are listed in Table 2. Among the sand fractions, total, medium, and fine sands had the highest mean values. Bulk density exhibited lower variability than the other attributes, which is consistent with the findings of Siqueira et al. (2008). The statistical distributions of the input and output variables are shown in Fig. 3.

Relationship between variables

Figure 4 presents the magnitude and direction of the linear relationships between the studied variables. Based on Pearson's coefficient classification presented by Ratner (2009), a negative correlation was observed between soil density and smaller particles: moderate ($0.5 \leq |r| < 0.7$) for clay (CL) and weak ($0.3 \leq |r| < 0.5$) for silt. In contrast, the density and larger fractions showed a positive correlation, which was weak for coarse sand (CO) and medium sand (ME), and negligible ($|r| < 0.3$) for very coarse sand (VC) and fine sand (FI). This result is consistent with Pádua et al. (2015), who reported a correlation of 0.46 between sand content

Table 2 Descriptive statistics of the physical attributes of soils

Attribute		Samples	Mean	Minimum	Maximum	SD	CV(%)
Very coarse	g kg^{-1}	368	26.818	5.98	75.15	17.129	63.87
Coarse	g kg^{-1}	368	54.078	20.08	135.59	25.448	47.06
Medium	g kg^{-1}	368	263.583	63.69	501.59	100.777	38.23
Fine	g kg^{-1}	368	259.393	111.10	496.30	68.227	26.30
Very fine	g kg^{-1}	368	151.085	48.80	265.00	42.609	28.20
Silt	g kg^{-1}	368	63.344	0.35	312.74	46.738	73.78
Clay	g kg^{-1}	368	181.709	39.00	407.80	83.777	46.11
Bulk density	kg dm^{-3}	368	1.506	1.33	1.77	0.086	5.70

SD standard deviation, *CV* coefficient of variation

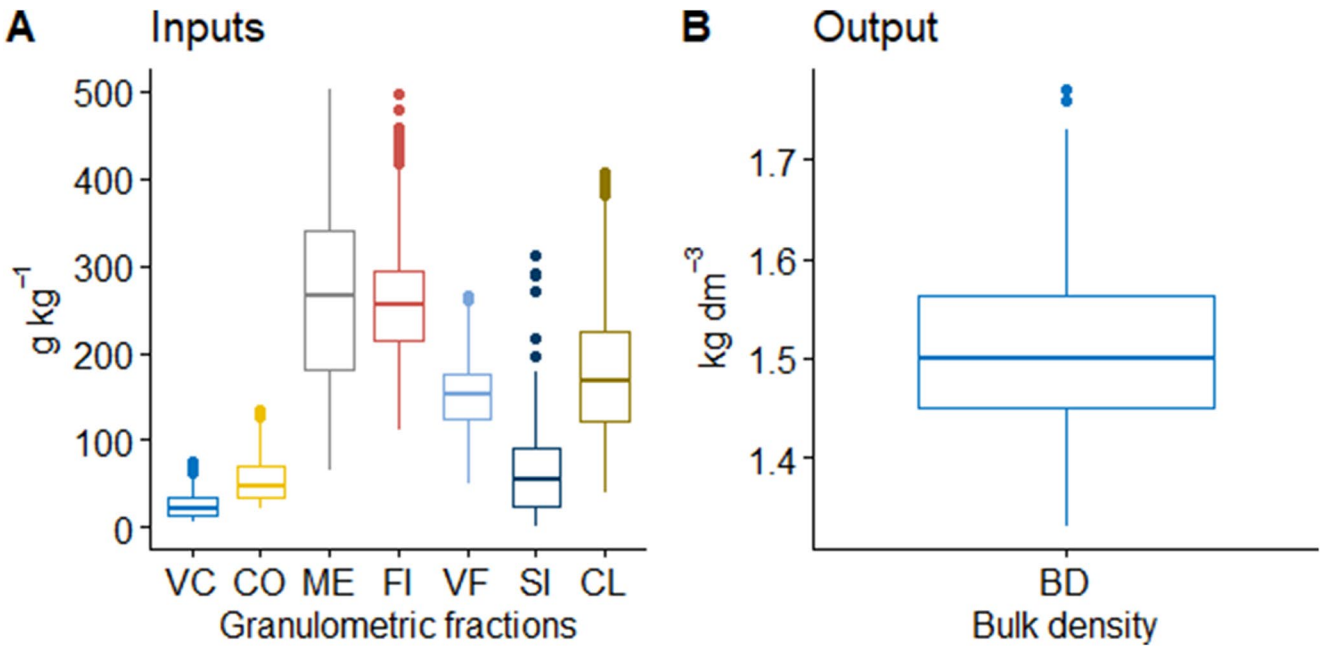

Fig. 3 Box plots of (**a**) particle size composition and (**b**) bulk density. Note: Very coarse (VC); coarse (CO); medium (ME); fine (FI); and very fine (VF) sand; clay (CL); silt (SI); and bulk density (BD)

and bulk density at depth of the 0–100 cm. According to the authors, the comparison consisted of a straightforward Pearson correlation of the measured bulk density and sand percentage data across all depths combined.

Regarding smaller particles, silt and clay fractions contribute favorably to reduce bulk density because of their intrinsic physical and chemical characteristics that promote a porous and less compact structure than sand-dominated soils (Li and Tang 2022). Owing to the greater surface area of clay per unit volume, its surface electrical charges facilitate bonding between clay and other soil fractions, such as silt and organic matter, forming stable aggregates (Luciano et al. 2012). These structures form intra- and inter-aggregate pore spaces, contributing to increased total porosity and, consequently, density reduction (Masís-Meléndez et al. 2015). This negative correlation was also observed by Barros and Fearnside (2015), Luciano et al. (2012), Silva et al. (2018), and Souza et al. (2016).

Silt, although not as reactive as clay, contributes to space filling and can participate in microaggregate formation, particularly in the presence of clay and organic matter (Li and Tang 2022). Pádua et al. (2015) further emphasized that the presence of silt and clay favors the development of granular and blocky structures, which are more porous than the simple grains typical of sandy soils.

In contrast, the positive linear relationship between sand fraction and soil bulk density can be explained by their more regular granular form, which contributes to compaction, especially when there is a mixture of different sand fractions, in which filling of interstitial spaces is preponderant (Abrahão et al. 1998). Furthermore, compared to silt and clay, sand particles are chemically less reactive and possess relatively smaller specific surface areas, resulting in a lower tendency to form stable aggregates, which are crucial for maintaining soil macroporosity (Pádua et al. 2015).

Neuro-fuzzy model

The selected model possessed an ANFIS adaptive network structure, according to the specifications indicated in Table 3. The learning technique was hybrid, consisting of the backpropagation of parameters associated with membership function inputs and least squares estimation of parameters associated with these function outputs. These functions were triangular, with coefficients a, b, and c (Eq. 12), in which each crisp number in the input range was assigned a degree of membership between 0 and 1, indicating its belonging to a particular fuzzy set (Table 4).

$$f(x;a,b,c) = \begin{cases} 0, & x \leq a \; and \; c \leq x. \\ \frac{x-a}{b-a}, & a \leq x \leq b. \\ \frac{c-x}{c-b}, & b \leq x \leq c. \end{cases} \tag{12}$$

The ANFIS architecture employed triangular membership functions for each of the seven input variables (VC, CO, ME, FI, VF, SI, and CL), giving each input two linguistic terms (IF-part in Fig. 2) that are defined by three coefficients of the triangular membership function (MF). With seven inputs and two MFs, the total number of adaptive (nonlinear) nodes was $7 \times 2 \times 3 = 42$. The total number of nodes in

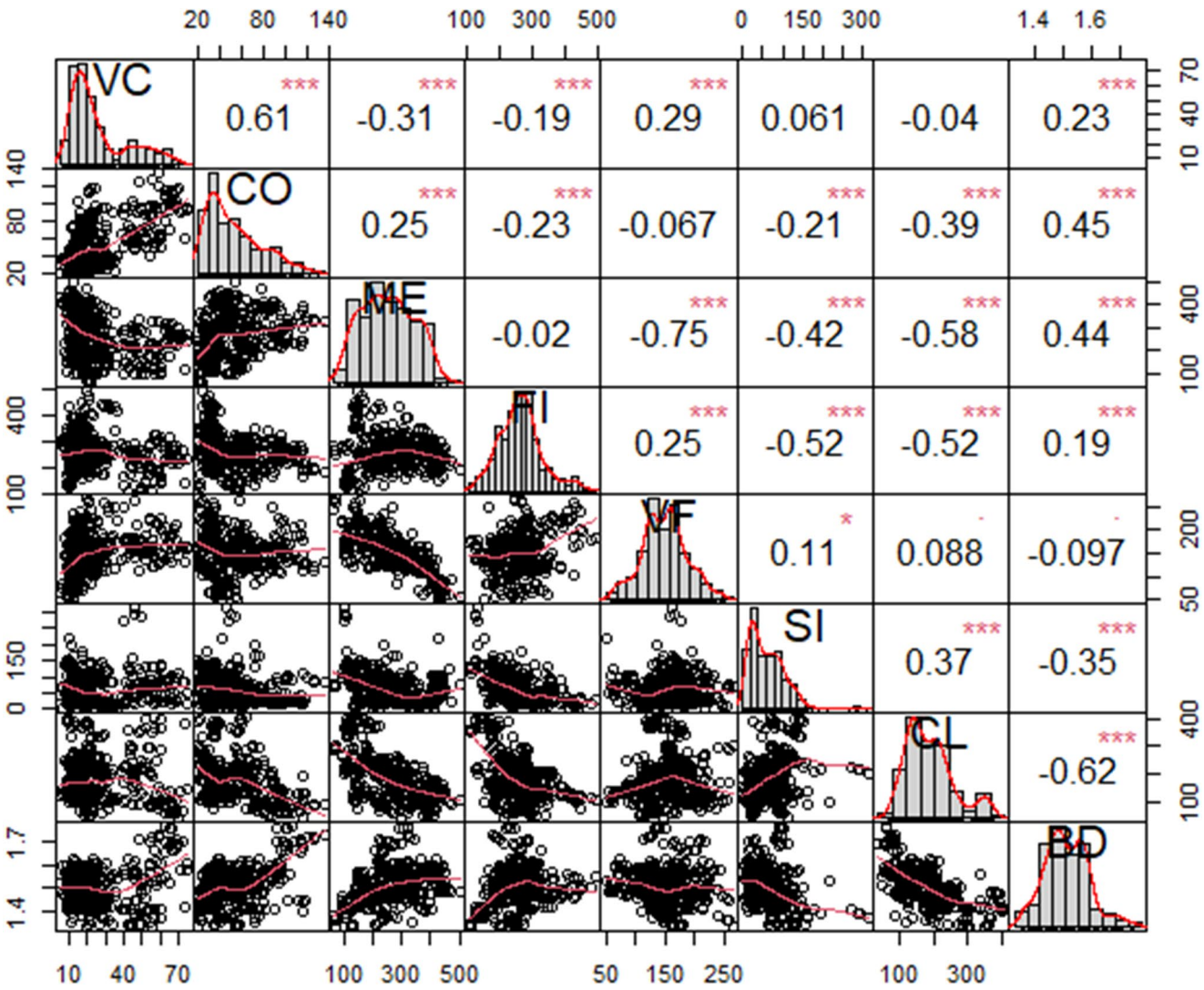

Fig. 4 Relationship between the studied variables. Significance level and symbol convention: 0–0.001.001 (***), 0.001–0.01 (**), and 0.01–0.05 (*)

Table 3 ANFIS parameters

Parameter	Specification
Membership function	Triangular
Number of nodes	294
Number of linear parameters	1024
Number of non-linear parameters	42
Epochs	1000
Number of training data pairs	276
Number of checking data pairs	92
Defuzzification output	Affine function
Error tolerance	Error = 0
Number of fuzzy rules	128
Optimization method	Hybrid

the network (including the input, fuzzy, and output layers) comprised 294 ones. The figure was obtained by adding the 7 input nodes, 14 MF nodes, 128 rule nodes (2^7 possible IF-THEN combinations), and 2 output-layer nodes (error back-propagation and defuzzification).

The remaining ANFIS settings were derived directly from the model-building process. Each of the 128 fuzzy rules contributed a linear consequent ($y = \sum a_i x_i + b$), providing eight coefficients per rule (seven inputs and one bias), yielding $128 \times 8 = 1024$ linear parameters. Training was performed for 1000 epochs on 276 randomly selected samples (70% of the 257 observations), while 56 and 55 samples (approximately 15%) served as validation and test

Table 4 Input range and triangular function coefficients

		Membership function (MF) coefficients (Fuzzification)					
		1th MF			2nd MF		
Input Range		a	b	c	a	b	c
VC	[8.055, 74.95]	−58.84	5.60	74.27	12.11	72.32	141.84
CO	[20.07, 127.4]	−87.28	19.44	127.20	24.00	126.79	234.79
ME	[66.19, 501.6]	−369.20	66.13	501.10	66.29	501.52	936.99
FI	[111.3, 496.3]	−273.70	111.10	495.80	111.94	496.07	881.26
VF	[48.8, 265]	−167.40	48.91	265.00	49.04	265.15	481.28
SI	[39, 403]	−325.00	38.88	403.00	40.40	402.88	767.00
CL	[0.35, 312.7]	−312.00	−0.13	312.70	0.34	312.25	625.13

sets respectively. The defuzzification method was an affine (linear) function, and the error tolerance was set to zero (exact convergence criterion).

Model performance

Figure 5 illustrates the relationship between the soil bulk density values predicted from the neuro-fuzzy simulation (P) and the experimentally observed values (O), with the ideal prediction represented by the diagonal line (P=O). The model adequately fits the testing data, obtaining a correlation coefficient of 84% (P-value<0.05) (Fig. 5).

The accuracy of bulk density prediction after model test is summarized by the statistics presented in Table 5. A bias of −0.0009 indicates the absence of the model's tendency to make predictions systematically above or below the observed density values. The low magnitude of the errors was also confirmed by the AME, MSE, RMSE, RMSLE, and MAPE. The NSE, commonly used in model performance evaluation in hydrology (Tongal and Booij 2018), was 69.5%, indicating good neuro-fuzzy model performance.

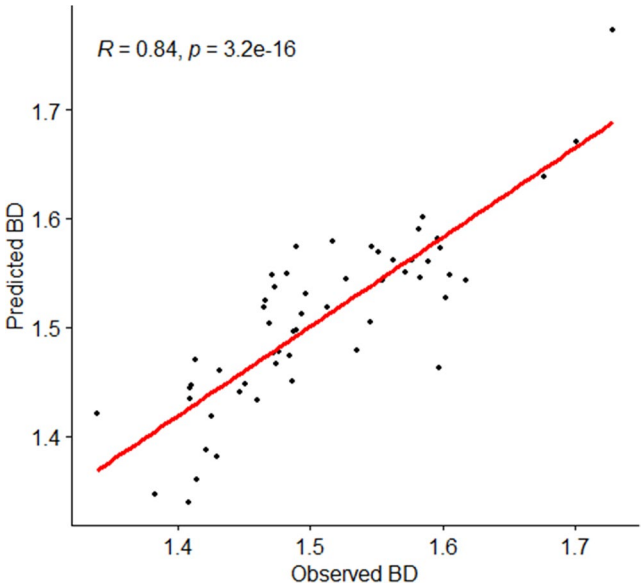

Fig. 5 Relationship between predicted and observed BD values after testing the model

The coefficient of determination R^2 was greater than some pedotransfer models available in the literature and developed from data from Brazilian soil samples, as shown in Tabel 5. It is important to note that each reported R^2 value reflects the individual study's measure and cannot be used to compare the performance of the models, because they involve different numbers of predictor variables and other factors that differentiate the data.

According to Table 6, most of these functions were obtained using multiple linear regression (MLR). In these studies, each independent variable was isolated and held constant while the others were systematically varied to observe their effects on soil density. In some analyses, sets of specific soil particle size compositions required dedicated models to enhance prediction accuracy (Bernoux et al. 1998; Pádua et al. 2015; Chen et al. 2024; Arbor et al. 2024; Gu et al. 2025; Bashir et al. 2025).

The studies selected for comparison (Table 6) developed pedotransfer functions to predict soil bulk density using readily measured soil properties, and typically incorporated organic carbon, clay content, and pH as the most influential predictors. A recurring finding was that stratifying the dataset by soil taxonomy markedly improved model performance. For example, Pádua et al. (2015) achieved an adjusted R^2 of 0.85 for Oxisols when data were grouped by order, and Barros and Fearnside (2015) reported an R^2 of 0.73 for locally calibrated models versus poorer fits of generic equations. Bernoux et al. (1988) obtained R^2 values up to 0.79 for A horizons in Oxisols but lower for Podzolicos, highlighting taxonomic dependence.

Table 5 Model error statistics

Error statistics	Value
R^2	0.7120
Bias	−0.0009
Mean absolute error	0.0355
Mean squared error	0.0020
Root mean squared error	0.0445
Root mean squared logarithmic error	0.0178
Mean absolute percentage error	0.0236
Nash-Sutcliffe efficiency coefficient	0.6946

Table 6 Comparison of bulk density prediction models

Authors	Predictor variables [1]	Method	R^2
Pádua et al. 2015	TS, CL, SOC, CEC, SI	MLR[2]	0,51 − 0,85[3]
Bernoux et al. 1998	SOC, TS, CL, pH	MLR	0,37 − 0,79
Barros and Fearnside 2015	CL, pH	MLR	0,73
This work	VC, CO, ME, FI, VF, SI, CL	*Neuro-fuzzy*	0,71
Benites et al. 2007	CL, SOC, SB	MLR	0,66
Souza et al. 2016	CL, SOC, CEC, pH, SB, SI[4]	*Random Forest*	0,51
		MLR	0.47
Beutler et al. 2017	CL, SOC	MLR	0,47

(1) TS: total sand; CL: clay content; SOC: soil organic carbon; CEC: cation-exchange capacity; SB: sum of bases; VC: very coarse sand; CO: coarse sand; ME: medium sand; FI: fine sand; VF: very fine sand; SI: silt particles; and bulk density (BD)

(2) MLR: multiple linear regression

(3) Adjusted coefficient of determination

(4) The models include the environmental variables

The simplified model developed by Benites et al. (2007) explained 66% of the variance and exhibited the lowest bias among several benchmark equations, while Beutler et al. (2017) found that a SOC and CL function still retained reasonable predictive power and that models developed for mineral soils performed poorly on high-organic-matter horizons. Collectively, these studies prove that horizon-specific calibration and data stratification are key to achieving accurate bulk density predictions. This demonstrates that the neuro-fuzzy methodology provides a more reliable and versatile tool for estimating sandy soil bulk density than conventional models.

Response surface

The analysis of the neuro-fuzzy model response surface allowed the examination of complex relationships between particle size fractions and bulk density (BD). Figure 6 illustrates the simultaneous effects of fine sand (FI) and clay (CL) on soil density. Both the increase in the fine sand and the clay contents contributed to reducing the response variable.

Although the fine sand fraction is commonly associated with an increased density by filling the space between larger particles and, consequently, forming a denser arrangement (Luciano et al. 2012), Braida et al. (2006) described a possible reduction in this variable attributed to fine sand. In this case, the particle density of this fraction is intrinsically lower than that of other mineral constituents, contributing to the relative reduction in the bulk density. This particularity was also observed by Wang et al. (2022), who investigated

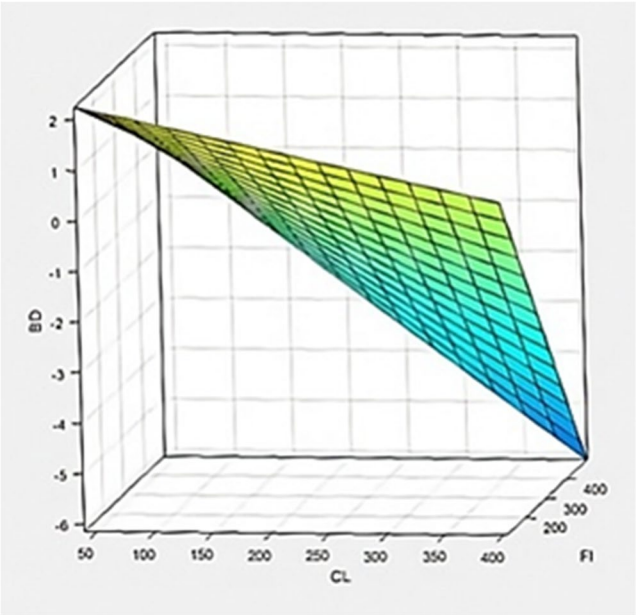

Fig. 6 Effect of clay (CL) and fine sand (FI) on bulk density (BD)

specific sedimentary environments in five regions with different deposition factors.

Furthermore, Cunha et al. (2010) reinforced the importance of sand fractions in soils representative of the present study area, which are predominantly composed of quartz. For example, in Quartzarenic Neosols (Entisol – Quartzipsamment), which possess 95% or more quartz in coarse (CO) and fine (FI) sand fractions, the coarse texture and high porosity generally imply, according to the authors, a lower bulk density.

In Figs. 7 and 8, an increase in the clay content (CL) is observed to lead to a reduction in bulk density. Han et al.

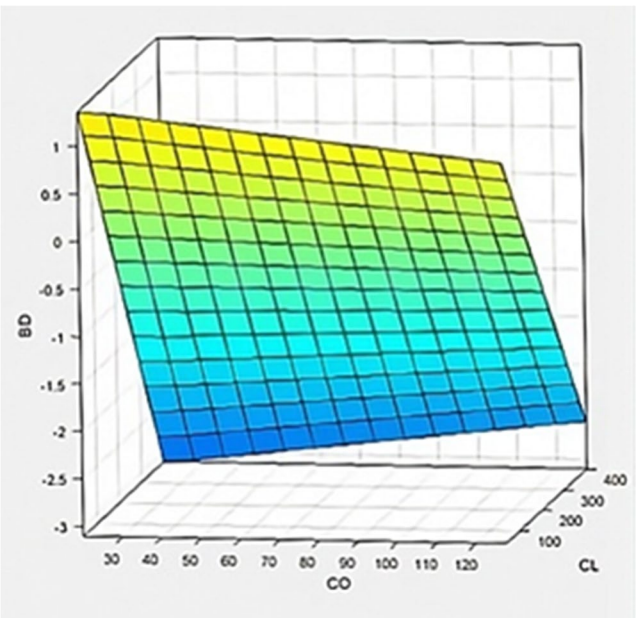

Fig. 7 Effect of coarse sand (CO) and clay (CL) on bulk density (BD)

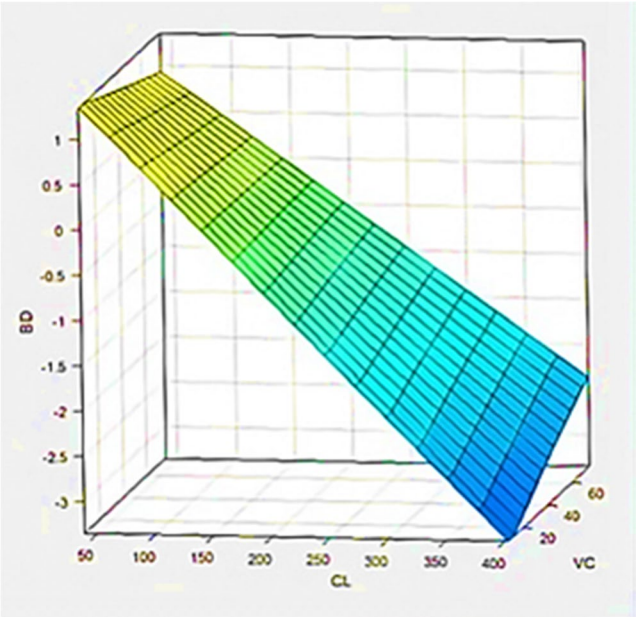

Fig. 8 Effect of clay (CL) and very coarse sand (VC) on bulk density (BD)

(2021) highlighted the capacity of clay to form aggregates, the mechanism by which it influences soil structure and reduces its density. In semi-arid regions, where organic matter content may be naturally low, clay's capacity to stabilize it is even more critical for maintaining good structure and lower density (Araujo Filho et al. 2022).

Finally, the model revealed another complex relationship, as shown in Fig. 9. The figure shows the distinct influence of very fine sand (VF) as the coarse sand fraction (CO) varies.

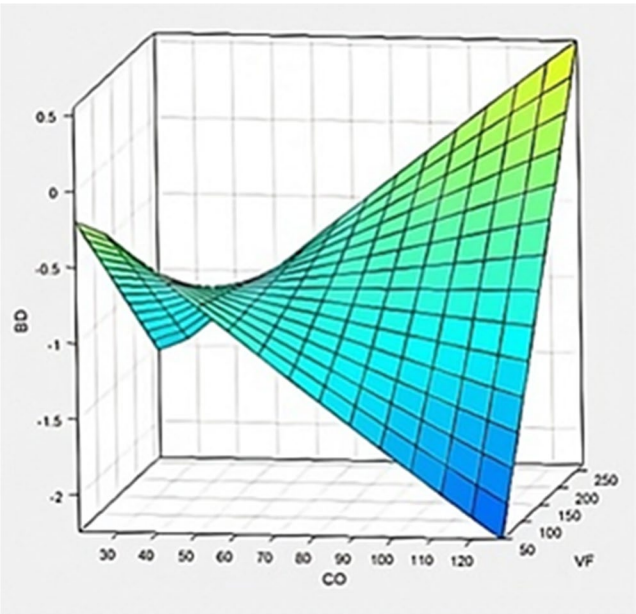

Fig. 9 Effect of coarse sand (CO) and very fine sand (VF) on bulk density (BD)

At lower CO levels, an increase in the VF reduces the BD. This is because VF particles have intrinsically lower particle densities than other mineral constituents, as described by Braida et al. (2006), an effect also observed by Wang et al. (2022) in different sedimentary environments.

In contrast, at higher CO contents, the relationship reverses, and an increasing VF contributes to BD increase. Here, VF particles fill the interstitial spaces between larger coarse sand particles, creating a denser packing arrangement as explained by Luciano et al. (2012). Cunha et al. (2010) reinforce that in quartz-dominated soils, coarse texture generally implies lower BD, but this can be altered by such particle interactions. This reversal demonstrates that the effect of VF fundamentally depends on the CO framework present, highlighting the need to consider particle size distribution interactions rather than isolated fractions when predicting soil physical properties.

Conclusion

The neuro-fuzzy model explained the relationship between soil particle size composition and bulk density, presenting R^2 of 71%. Furthermore, satisfactory performance was observed in density predictions in soils according to their granulometric fractions, demonstrating resilience in generalization capacity across different areas and soil collection horizons. It is important to emphasize that the neuro-fuzzy model did not require data stratification and can be applied in estimating sandy soil density. Moreover, the response-surface leads to the conclusion that increasing the fine-sand (FI) and clay (CL) fractions consistently lowers bulk density, indicating that finer particles promote a more porous, less compact soil structure, and the influence of very-fine sand (VF) on bulk density depends on the amount of coarse sand (CO): at low CO levels, more VF reduces bulk density, whereas at higher CO levels, additional VF actually raises bulk density, revealing a conditional interaction between these sand fractions.

While the model presents itself as a practical and rapid alternative to laboratory analysis for obtaining soil density values in sandy soils, some questions remain for broader agricultural application. A primary research direction is whether comparable neuro-fuzzy frameworks can be calibrated for loams, clays, or more neutral soil textures, given that studies evaluating bulk density prediction with intelligent models are still scarce. In addition, comparative assessments with alternative machine-learning techniques and larger, multi-regional datasets are also needed to gauge generalizability. Finally, integrating remote-sensing-derived texture maps and real-time mechanization metrics could extend the tool into a versatile decision-support system for sustainable soil management across diverse soil types.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12145-026-02077-y.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). The data used in the article originated from the eolic energy project "Eólicas Casa Nova", a partnership between Embrapa and AXIA Energy.

Author contributions Dian Lourençoni : Conceptualization, Methodology, Validation. Alessandra Salviano : Investigation, Resources, Writing - Review & Editing. Nelci Olszevski : Investigation, Resources, Writing - Review & Editing, Supervision. Janielle Pereira : Investigation, Resources. Edmo Cavalcante : Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization.

Funding The Article Processing Charge (APC) for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). The data used in the article originated from the eolic energy project "Eólicas Casa Nova", a partnership between Embrapa and AXIA Energy.

Data availability The data were deposited into the Figshare repository and are available at the following URL: https://doi.org/10.6084/m9.figshare.30522689.

Declarations

Competing interests The authors declare no competing interests.

References

Abrahão WAP, Costa LM, Mello JWV, Neves JCL (1998) Distribuição de freqüência de Tamanho Da fração Areia e compacidade relativa de solos desenvolvidos de sedimentos do grupo geológico Barreiras. Rev Bras Cienc Solo 22:1–9. https://doi.org/10.1590/S0100-06831998000100001

Akgun A, Sezer EA, Nefeslioglu HA et al (2012) An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput Geosci 38:23–34. https://doi.org/10.1016/j.cageo.2011.04.012

Araujo Filho JC, Marques FA, Amaral AJ et al (2022) Solos do semiárido: características e estoque de Carbono. In: Giongo V, Angelotti F (eds) Agricultura de Baixa Emissão de Carbono Em regiões semiáridas: experiência Brasileira. Embrapa, Brasília, pp 93–112

Arbor A, Schmidt M, Zhang J et al (2024) Filling the gaps in soil data: a multi-model framework for addressing data gaps using pedotransfer functions and machine-learning with uncertainty estimates to estimate bulk density. CATENA 245:108310. https://doi.org/10.1016/j.catena.2024.108310

Arcoverde SNS, Salviano AM, Olszevski N et al (2015) Qualidade física de solos em uso agrícola na região semiárida do estado da Bahia. Rev Bras Cienc Solo 39:1473–1482. https://doi.org/10.1590/01000683rbcs20140282

Barbosa TCS, Costa NMGB, Santos DB et al (2020) Qualidade física do solo Em áreas Sob Manejo agroecológico e convencional. Brazilian J Dev 6:48899–48909. https://doi.org/10.34117/bjdv6n7-511

Barros HS, Fearnside PM (2015) Pedo-transfer functions for estimating soil bulk density in central Amazonia. Rev Bras Cienc Solo 39:397–407. https://doi.org/10.1590/01000683rbcs20140358

Bashir O, Bangroo SA, Amjid S et al (2025) Soil bulk density estimation by a novel pedotransfer function tailored on hilly terrains. Earth Syst Environ 9:1615–1633. https://doi.org/10.1007/s41748-025-00663-6

Bengough AG, Bransby MF, Hans J et al (2006) Root responses to soil physical conditions; growth dynamics from field to cell. J Exp Bot 57:437–447. https://doi.org/10.1093/jxb/erj003

Benites VM, Machado PLOA, Fidalgo ECC et al (2007) Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma 139:90–97. https://doi.org/10.1016/j.geoderma.2007.01.005

Bernoux M, Cerri C, Arrouays D et al (1998) Bulk densities of brazilian amazon soils related to other soil properties. Soil Sci Soc Am J 62:743–749. https://doi.org/10.2136/sssaj1998.03615995006200030029x

Beucher A, Fröjdö S, Österholm P et al (2014) Fuzzy logic for acid sulfate soil mapping: application to the southern part of the Finnish coastal areas. Geoderma 226:21–30. https://doi.org/10.1016/j.geoderma.2014.03.004

Beutler SJ, Pereira MG, Tassinari WS et al (2017) Bulk density prediction for histosols and soil horizons with high organic matter content. Rev Bras Cienc Solo 41. https://doi.org/10.1590/18069657rbcs20160158

Braida JA, Reichert JM, Veiga M, Reinert DJ (2006) Resíduos Vegetais Na superfície e Carbono orgânico do solo e Suas relações com a densidade máxima obtida no ensaio proctor. Rev Bras Cienc Solo 30:605–614. https://doi.org/10.1590/S0100-06832006000400001

Branco RNC, Poça RR, Gomes RCC et al (2021) Sistema fuzzy Para Tomada de decisão Acerca Da qualidade do solo Na Amazônia Brasileira / Fuzzy system for decision making about soil quality in the Brazilian Amazon. Brazilian J Dev 7:86269–86281. https://doi.org/10.34117/bjdv7n8-705

Chen S, Chen Z, Zhang X et al (2024) European topsoil bulk density and organic carbon stock database (0–20 cm) using machine-learning-based pedotransfer functions. Earth Syst Sci Data 16:2367–2383. https://doi.org/10.5194/essd-16-2367-2024

Cunha JPAR, Vieira LB, Magalhães AC (2002) Resistência mecânica do solo à penetração Sob diferentes densidades e Teores de água. Engenharia Agric 10:1–7

Cunha TJF, Petrere VG, Silva ADJ et al (2010) Principais solos do Semiárido tropical brasileiro: Caracterização, potencialidades, limitações, fertilidade e Manejo. In: Sá IB, Silva PCG (eds) Semiárido brasileiro: pesquisa, desenvolvimento e inovação. Embrapa Semiárido, Petrolina, pp 50–87

Forster M, Ugarte C, Lamandé M, Faucon M (2020) Relationships between root traits and soil physical properties after field traffic from the perspective of soil compaction mitigation. Agron 10:1697. https://doi.org/10.3390/agronomy10111697

Godinho EZ, Gasparotto HV, Caneppele FL (2022) Lógica fuzzy Na agricultura. Cadernos De Educação Tecnologia E Sociedade 15:126–139. https://doi.org/10.14571/brajets.v15.n1.126-139

Gomes Junior WL, Barbosa FT, Melo HF (2022) Root and shoot development in winter crops in soils of different textures and degrees of compaction. Revista Ciência Agronômica 54. https://doi.org/10.5935/1806-6690.20230020

Gu J, Ye M, Song X et al (2025) Bulk density datasets using Pedotransfer functions in the Qinghai-Tibetan plateau. Sci Data 12:649. https://doi.org/10.1038/s41597-025-04983-0

Halder B, Ahmed MM, Amagasa T et al (2022) Missing information in imbalanced data stream: fuzzy adaptive imputation approach. Appl Intell 52:5561–5583. https://doi.org/10.1007/s10489-021-02741-4

Hamza MA, Anderson WK (2005) Soil compaction in cropping systems. Soil Tillage Res 82:121–145. https://doi.org/10.1016/j.still.2004.08.009

Han Y, Wang Q (2023) Feature and mechanism analysis of dispersive soil disintegration impacted by soil water content, density, and salinity. Eur J Soil Sci. https://doi.org/10.1111/ejss.13353

Han Z, Li J, Li Y et al (2021) Assessment of the size selectivity of eroded sediment in a partially saturated sandy loam soil using scouring experiments. CATENA 201:105234. https://doi.org/10.1016/j.catena.2021.105234

Herrick JE (2000) Soil quality: an indicator of sustainable land management? Appl Soil Ecol 15:75–83. https://doi.org/10.1016/S0929-1393(00)00073-1

Horn R, Domżżał H, Słowińska-Jurkiewicz A, Ouwerkerk C (1995) Soil compaction processes and their effects on the structure of arable soils and the environment. Soil Tillage Res 35:23–36. https://doi.org/10.1016/0167-1987(95)00479-C

Jensen TA, Antille DL, Tullberg JN (2025) Improving on-farm energy use efficiency by optimizing machinery operations and management: a review. Agric Res 14:15–33. https://doi.org/10.1007/s40003-024-00824-5

Kuncoro PH, Koga K, Satta N, Muto Y (2014) A study on the effect of compaction on transport properties of soil gas and water I: relative gas diffusivity, air permeability, and saturated hydraulic conductivity. Soil Tillage Res 143:172–179. https://doi.org/10.1016/j.still.2014.02.006

Labelle ER, Kammermeier M (2019) Above- and belowground growth response of *Picea abies* seedlings exposed to varying levels of soil relative bulk density. Eur J Res 138:705–722. https://doi.org/10.1007/s10342-019-01201-6

Lins ACSS, Lourençoni D, Yanagi Júnior T et al (2021) Neuro-fuzzy modeling of eyeball and crest temperatures in egg-laying hens. Engenharia Agrícola 41:34–38. https://doi.org/10.1590/1809-4430-eng.agric.v41n1p34-38/2021

Li T, Tang X (2022) Influences of low fines content and fines mixing ratio on the undrained static shear strength of sand–silt–clay mixtures. Eur J Environ Civ Eng 26:3706–3728. https://doi.org/10.1080/19648189.2020.1813206

Luciano RV, Albuquerque JA, Costa A et al (2012) Atributos físicos relacionados à compactação de solos sob vegetação nativa em região de altitude no Sul do Brasil. Rev Bras Cienc Solo 36:1733–1744. https://doi.org/10.1590/S0100-06832012000600007

Masís-Meléndez F, Jonge LW, Deepagoda TKKC et al (2015) Effects of soil bulk density on gas transport parameters and pore-network properties across a sandy field site. Vadose Zone J 14:1–12. https://doi.org/10.2136/vzj2014.09.0128

Matsinhe DZ, Pereira SB, Oliveira RA et al (2019) Use of PVC pipes to determine bulk density for irrigation management. Revista Ciência Agronômica 50. https://doi.org/10.5935/1806-6690.20190004

McPhee JE, Antille DL, Tullberg JN et al (2020) Managing soil compaction: a choice of low-mass autonomous vehicles or controlled traffic? Biosyst Eng 195:227–241. https://doi.org/10.1016/j.biosystemseng.2020.05.006

Mendel JM (2024) Sources of uncertainty and membership functions. Explainable Uncertain Rule-Based Fuzzy Systems. Springer International Publishing, Cham, pp 219–235

Mileusnić ZI, Saljnikov E, Radojević RL, Petrović DV (2022) Soil compaction due to agricultural machinery impact. J Terramech 100:51–60. https://doi.org/10.1016/j.jterra.2021.12.002

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — A discussion of principles. J Hydrol (Amst) 10:282–290. https://www.sciencedirect.com/science/article/abs/pii/0022169470902556

Novák V (2006) Which logic is the real fuzzy logic? Fuzzy Sets Syst 157:635–641. https://doi.org/10.1016/j.fss.2005.10.010

Omomule TG, Ajayi OO, Orogun AO (2020) Fuzzy prediction and pattern analysis of poultry egg production. Comput Electron Agric 171:105301. https://doi.org/10.1016/j.compag.2020.105301

Osna T, Sezer EA, Akgun A (2014) GeoFIS: an integrated tool for the assessment of landslide susceptibility. Comput Geosci 66:20–30. https://doi.org/10.1016/j.cageo.2013.12.016

Pádua EJ, Guerra AR, Zinn YL (2015) Modelagem Da densidade do solo Em profundidade Sob vegetação nativa Em Minas Gerais. Rev Bras Cienc Solo 39:725–736. https://doi.org/10.1590/01000683rbcs20140028

Queiroz LP, Cardoso D, Fernandes MF, Moro MF (2017) Diversity and evolution of flowering plants of the Caatinga domain. In: Silva JMC, Leal IR, Tabarelli M (eds) Caatinga: the largest tropical dry forest region in South America. Springer International Publishing, Cham

Raper RL (2005) Agricultural traffic impacts on soil. J Terramech 42:259–280. https://doi.org/10.1016/j.jterra.2004.10.010

Ratner B (2009) The correlation coefficient: its values range between +1/–1, or do they? J Target Meas Anal Mark 17:139–142. https://doi.org/10.1057/jt.2009.5

Rima L, Zekari M, Rym TH et al (2024) Dimensional analysis based prediction model for fuel consumption in the Deutz agrotrac tractor. https://doi.org/10.18805/ag.DF-561. Agricultural Science Digest - A Research Journal

Robinson DA, Thomas A, Reinsch S et al (2022) Analytical modelling of soil porosity and bulk density across the soil organic matter and land-use continuum. Sci Rep 12:7085. https://doi.org/10.1038/s41598-022-11099-7

Rodriguez L, Bastidas M, Silva M et al (2025) Comparative assessment of soil bulk density measurements using core metal ring and power probe methods in acidic soils of Colombian pasturelands. Front Soil Sci 5:1527807. https://doi.org/10.3389/fsoil.2025.1527807

Santos IEA, Okita WM, Lourençoni D et al (2024) Neuro-fuzzy modeling of pulp temperature in rapid cooling chamber. J Food Sci Technol. https://doi.org/10.1007/s13197-024-06109-7

Silva AC, Portela JC, Batista RO et al (2018) Soil water retention in the semiarid region of Brazil. J Agric Sci 10:105. https://doi.org/10.5539/jas.v10n9p105

Siqueira GM, Vieira SR, Ceddia MB (2008) Variabilidade de Atributos físicos do solo determinados Por métodos Diversos. Bragantia 67:203–211. https://doi.org/10.1590/S0006-87052008000100025

Souza E, Fernandes Filho EI, Schaefer CEGR et al (2016) Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. Sci Agric 73:525–534. https://doi.org/10.1590/0103-9016-2015-0485

Sturm N, Wilson KR, Neely HL et al (2025) Informing soil compaction research priorities with farmer focus groups in the United States. Soil Secur 20:100200. https://doi.org/10.1016/j.soisec.2025.100200

Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern 15:116–132. https://doi.org/10.1109/TSMC.1985.6313399

Tedeschi LO (2006) Assessment of the adequacy of mathematical models. Agric Syst 89:225–247. https://doi.org/10.1016/j.agsy.2005.11.004

Teixeira PC, Donagemma GK, Fontana A, Teixeira WG (2017) Manual de métodos de análise de solo, 3rd edn. Embrapa, Brasília

Tongal H, Booij MJ (2018) Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. J Hydrol (Amst) 564:266–282. https://doi.org/10.1016/j.jhydrol.2018.07.004

USDA (2023) Precision agriculture and soil management practices. In: Economic Research Service Report. https://www.ers.usda.gov. Accessed 10 Sep 2025

Wang Y, He Y, Zhan J, Li Z (2022) Identification of soil particle size distribution in different sedimentary environments at river basin scale by fractal dimension. Sci Rep 12:10960. https://doi.org/10.1038/s41598-022-15141-6

Zhang H, Wang C, Chen Z et al (2022) Performance comparison of different particle size distribution models in the prediction of soil particle size characteristics. Land 11:2068. https://doi.org/10.3390/land11112068

Zheng G, Jiao C, Xie X et al (2023) Pedotransfer functions for predicting bulk density of coastal soils in East China. Pedosphere 33:849–856. https://doi.org/10.1016/j.pedsph.2023.01.014

Zhu A, Yang L, Li B et al (2010) Construction of membership functions for predictive soil mapping under fuzzy logic. Geoderma 155:164–174. https://doi.org/10.1016/j.geoderma.2009.05.024