

Machine learning-based soil aggregation assessment under four scenarios in northwestern Iran

Parastoo Nazeri¹, Shamsollah Ayoubi^{1*}, Hossein Khademi¹, Farideh Abbaszadeh Afshar^{1,2}, Seyed Roohollah Mousavi³

¹Department of Soil Science, College of Agriculture, Isfahan University of Technology, 8415683111, Isfahan, Iran

²Department of Soil Science, Faculty of Agriculture, University of Jiroft, Jiroft 78671-61167, Iran

³Department of Soil Science, Faculty of Agriculture, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

Received February 19, 2024; accepted May 9, 2024

Abstract. Soil aggregate stability is crucial for maintaining the arrangement of solid particles and pore space in the soil, even under mechanical stresses. Traditional direct measurements of soil aggregate stability are time-consuming and expensive. This study aimed to spatially predict the soil aggregate stability indices, including the mean weight diameter of aggregates, the geometric mean diameter of aggregates, and the percentage of water stable aggregates, using five machine learning models and environmental covariates in the framework of digital soil mapping. A total of 100 samples were collected from the surface layer (0-15 cm) of soils in the Aji-Chai watershed, northwestern Iran, and their SAS indices were determined by standard laboratory methods. Four scenarios (S) were employed to evaluate the most influencing auxiliary variables, including (S₁): topographic attributes, (S₂): topographic attributes + remote sensing data, (S₃): S₂ + thematic maps (geology, land use/cover maps), and (S₄): S₃ + selected soil properties. Among the various machine learning models, the random forest showed exceptional performance and reduced uncertainty for S₄, compared to the other machine learning models and desired scenarios. The coefficient of determination, concordance correlation coefficient, and normalized root mean squared error values of the random forest model were 0.86, 0.87, and 31.42% for mean weight diameter; 0.80, 0.84, and 31.59% for geometric mean diameter; and 0.54, 0.68, and 20.75% for water stable aggregates, respectively. Additionally, properties such as soil organic matter and clay, followed by remote sensing data, demonstrated the highest relative importance when compared to the other covariates in predicting the soil aggregate stability indices. In conclusion, the random forest ML-based model seems to be able to accurately predict soil aggregate stability

indices at the watershed scale. The generated maps can serve as a valuable baseline for land use planning and decision-making. These findings contribute to the scientific understanding of soil physical quality indicators and their application in sustainable land management practices.

Keywords: soil aggregation, data mining, remote sensing, environmental covariates

1. INTRODUCTION

The sustainable management of soil is crucial for the well-being of our environment and the welfare of present and future generations (Wang *et al.*, 2017; Geisseler and Scow, 2014; World Health Organization, 2019). To achieve this, it is essential to have a deep understanding of the diverse characteristics of soil. One key aspect that plays a vital role in preserving and nurturing soil health is soil aggregate stability (SAS), which depends, to a great extent, on the combination of clay particles with organic matter. It determines physical processes, such as the self-organization, absorption, disposal, and storage of soil water (Dexter *et al.*, 2008; Resurreccion *et al.*, 2011; Farahani *et al.*, 2019). By focusing on maintaining SAS, we can effectively conserve and sustain the overall soil health in the world (Bhattacharyya *et al.*, 2021; Chaplot and Cooper, 2015; Hanke and Dick, 2017; Chahal and

Eerd, 2019; Samaei *et al.*, 2022). Aggregate stability indices serve as important indicators for evaluating soil deterioration and erodibility (Ye *et al.*, 2019). SAS indices play a crucial role in determining several key parameters related to soil quality and functioning. Parameters such as penetration resistance, susceptibility to soil erosion, fertility, aeration, soil carbon dynamics (including losses and sequestration), and biological activity are significantly influenced by SAS indices (Jastrow *et al.*, 1998; Deng *et al.*, 2014; Chaplot and Cooper, 2015). Furthermore, SAS directly impacts the storage and dynamics of soil organic matter (SOM) by physically protecting organic matter against oxidation and decomposition (Ayoubi *et al.*, 2020; Ayoubi *et al.*, 2012; Jastrow *et al.*, 1998; Zeraatpisheh and Khormali, 2012; Khosravani *et al.*, 2023). In fact, research conducted by Celik (2005) has demonstrated that SOM plays a direct role in the formation and stability of aggregates.

In recent years, the methodology of digital soil mapping (DSM) has gained recognition for its innovative and effective approach to illustrating the spatial distribution of soil properties. This technique has shown great promise in creating predictive maps of soil properties, as evidenced by studies (McBratney *et al.*, 2003; Wadoux *et al.*, 2020; Martin *et al.*, 2021). By leveraging patterns of soil-forming factors, the DSM framework can deduce soil information, overcoming the limitations of data in traditional mapping methods. Consequently, highly precise and detailed soil maps can be generated at a fine resolution (Liu *et al.*, 2020). Also, the number of environmental covariates used for DSM has rapidly increased due to the growing volume of RS data (Chen *et al.*, 2022). Previous studies have demonstrated that DSM techniques can accurately and efficiently describe the spatial distribution of soil properties (Ding *et al.*, 2016; Wu *et al.*, 2019; Song *et al.*, 2020; Mishra *et al.*, 2021). The majority of research in DSM has been focused on soil texture components (sand, silt, clay) as well as the percentage of soil organic carbon (SOC), nitrogen, phosphorus, and potassium. While there have been studies conducted by Zhao *et al.* (2022) on clay content mapping in Australia, Mousavi *et al.* (2022) on SOC in Iran, and Emadi *et al.* (2020) on potassium content in Brazil, which focused on analyzing specific soil properties, there is a lack of research on SAS indices in difficult-to-access areas.

Direct methods for assessing SAS can be cumbersome, labor-intensive, time-consuming, and expensive (Bhattacharyya *et al.*, 2021). As an alternative, indirect methods, also known as predictive soil maps by DSM approaches, offer a practical approach to estimate soil health indicators with reduced effort and cost while still providing reliable evaluation. These indirect methods can be valuable tools in soil research and management, allowing for efficient assessment of soil health without compromising accuracy (Martin *et al.*, 2021; Wadoux *et al.*, 2020; McBratney *et al.*, 2003). Importantly, the results of DSM are reproducible and able to quantify uncertainties (Arrouays *et al.*, 2020).

Generally, machine learning (ML) techniques are found to be successful in predicting soil parameters, in particular, geometric mean diameter (Besalatpour *et al.*, 2014), mean weight diameter (Bhattacharyya *et al.*, 2021), hydraulic conductivity (Rezaei *et al.*, 2023), soil organic carbon stocks (Rostaminia *et al.*, 2021), and field capacity moisture content (Yamaç *et al.*, 2020).

Remote sensing (RS) covariates, including surface reflectance of soils and vegetation indices, are widely utilized in DSM (McBratney *et al.*, 2003; Zhou *et al.*, 2021). Surface reflectance and band ratios are particularly effective in assessing SAS (Bouslihimi *et al.*, 2021a). In this regard, we can refer to the studies of Zeraatpisheh *et al.* (2021) on modeling the SAS using remote sensing covariates, soil properties, and topographic attributes and Wang *et al.* (2018), who emphasize applying RS data to predict soil properties. The combination of topographic and RS attributes has been confirmed in numerous studies in the field of soil properties prediction (Mashalaba *et al.*, 2020; Mousavi *et al.*, 2023).

In recent years, investigations carried out by Khosravani *et al.* (2023) and Zeraatpisheh *et al.* (2021) have focused on DSM of SAS. However, these studies have not addressed the issue of prediction uncertainty and have not explored the potential of utilizing time series remote sensing images for spatial modeling of SAS. Therefore, here we try to address this hypothesis by the incorporation of time series RS images and the utilization of advanced ML models which will enhance the accuracy and precision of spatial predictions of SAS indices in the Aji-Chai watershed in north-western Iran. However, it is important to note that there is a need for further investigation into properties such as mean weight diameter (*MWD*), geometric mean diameter (*GMD*), and percentage of water-stable aggregates (*WSA*). Understanding the relationships between these properties and representative soil-forming factors is crucial in the field of pedometric science and requires more attention, especially at the watershed scale. Therefore, the main objectives of this study were: (i) to predict the spatial distribution of SAS indices and uncertainty of prediction maps at the watershed scale, (ii) to compare different MLMs across various input scenarios of soil and environmental covariates, and (iii) to identify the most important environmental covariates that significantly influence prediction of SAS indices in the Aji-Chai watershed, East Azarbaijan Province, Iran.

2. MATERIALS AND METHODS

2.1. Study area

The Aji-Chai watershed is located in East Azarbaijan Province, located between 46°45' – 47°50' longitudes and 38°37' – 39°38' N latitudes. This watershed is located in the northwest of Iran (Fig. 1), with an area of around 1200 km², steepness ranging from 0 to more than 70%, with an altitude of 1800 m a.s.l. It is considered the largest

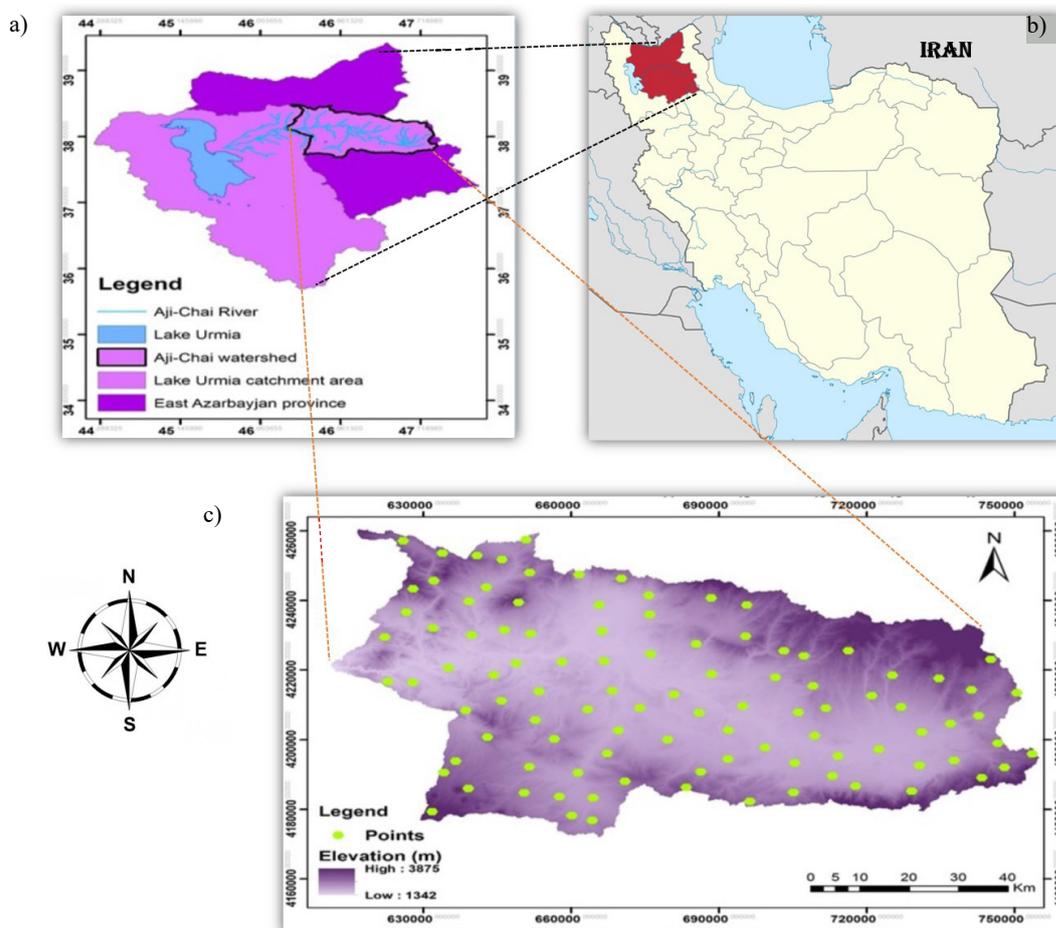


Fig. 1. Location of the: a) Aji-Chai watershed in East Azarbaijan province, b) East Azarbaijan province in northwestern Iran, and c) the spatial locations of the sampled soils in the studied watershed.

sub-basin of Urmia Lake after the Zarinah River sub-basin. The mean annual temperature and precipitation are 11°C and 320 mm, respectively. According to Dumarten's classification, this area has a semi-arid climate. The study area is predominantly characterized by pasture, which covers approximately 70% of the total area. Additionally, the watershed includes gardens, irrigated agricultural lands, and patches of bare land. Furthermore, based on the lithology map (1:100 000), igneous rocks, new alluviums, and marine deposits are the main parent materials contributing to soil formation in the study area.

2.2. Soil sampling and laboratory analysis

In the study area, sampling points were selected using the hypercube method based on their coordinates. A total of 100 sampling sites were chosen, and soil samples were collected from each site at a depth of 0-15 cm for physical and chemical analysis (Jiang, 2021). The collected soil samples were then transferred to the laboratory. As regards five soil physicochemical properties *e.g.* in a ratio of 1:2.5 soil: water suspension, soil pH and electrical conducti-

vity (EC) were determined (Page *et al.*, 1982). Also, clay was determined as soil particle size distribution with the pipette method (Gee and Bauder, 1986), soil organic matter (SOM) content was measured using the Walkley-Black method (Walkley and Black, 1934), and calcium carbonate equivalent (CCE) was measured using the titration method (Page *et al.*, 1982). The soil properties were selected based on expert opinion, and their correlation with SAS indices is displayed in Table 1.

100 grams of soil that passed through a 4.25 mm sieve was capillary wetted, and the wet sieving method was used to separate water-stable aggregates. The diameter of the sieves was 2-1, 0.5-0.25-0.1, and 0.05 mm, respectively. Then, the mean weight diameter of water-stable aggregates (*MWD*), the geometric mean diameter of water-stable aggregates (*GMD*), and the percentage of water-stable aggregates (% *WSA*) were determined. The percent of water-stable aggregates (% *WSA*) was calculated (Kemper and Rosenau, 1986) as follows:

$$\%WSA = \frac{w_{i(a+s)} - w_{i(s)}}{\sum_{i=1}^n w_t - \sum_{i=1}^n w_{i(s)}} 100, \quad (1)$$

where: \bar{X} is the arithmetic average of the size of aggregates on the i th sieve and W_i is the fraction of water-stable aggregates on the i th sieve, computed using:

$$W_i = \frac{w_{i(a+s)} - w_{i(s)}}{\sum_{i=1}^n w_{i(a+s)} - \sum_{i=1}^n w_{i(s)}}, \quad (2)$$

where: $W_i(a+s)$ is the dry weight of the particles on sieve i , $W_i(s)$ is the dry weight of sand or gravel on sieve i , W_t is the total dry weight of the soil, and n is the number of aggregate fractions. The MWD of water-stable aggregates was computed using Eq. (3) (Kemper and Rosenau, 1986):

$$MWD = \sum_{i=1}^n W_i \bar{X}. \quad (3)$$

Moreover, Eq. (4) was used to calculate the geometric mean diameter (GMD , mm) of aggregates:

$$GMD = \exp\left(\sum_{i=1}^n W_i \log \bar{X}\right). \quad (4)$$

2.3. Environmental covariates

Several environmental covariates were acquired from different sources to explain SAS indices. They include terrain attributes, RS data, and thematic maps (*i.e.* geology, land use maps). A preprocessed digital elevation model (DEM, 60 × 60 m resolution) from SRTM of the National Aeronautics and Space Administration (NASA) (Jarvis *et al.*, 2008) was downloaded to calculate terrain attributes. For RS indices, the median of satellite images of Landsat-8 operational land imager (OLI) and Sentinel-2 was used within the sampling time (July, 2022). Also, some soil variables were used as soil covariates for related modeling processes. In general, 120 covariates including vegetation indices were prepared from band ratios of RS satellite images and the first- and second-order DEM derivatives in SAGA (System for Automated Geoscientific Analysis) software (version 4.7).

2.3.1. Feature selection

Applying fewer covariates could benefit and improve the efficiency of the modeling process in DSM research (Brungard *et al.*, 2015, Yu *et al.*, 2024). The soil and environmental covariates were selected using two primary methods: variable importance analysis by the Random Forest algorithm and expert opinion. Subsequently, Pearson's correlation coefficient (r) was employed to assess the relationship between SAS indices and the soil properties chosen based on expert opinion (Table 1). In conclusion, 23 highly important and correlated covariates (refer to Table 2) were chosen for modeling. Six of the most important covariates are illustrated in Fig. 2 as representative of input covariates to the spatial modeling of SAS.

2.4. Spatial modeling of SAS

Five popular MLMs were employed, including random forest (RF), k-Nearest Neighbor (k-NN), support vector regression (SVR), artificial neural networks (ANNs), and cubist (CB) for modeling and mapping of SAS indices under four scenarios (Table 3). All the MLMs were implemented using specific packages of the R statistical software version (4.2.1).

2.4.1. Random forest (RF)

Random forest (RF) is an extended version of the regression and classification tree models (Breiman, 2001). It is one of the most frequently used ML models in DSM. The RF output is the mean of all regression trees, and the ensemble model is built by averaging multiple models based on different bootstrap sample datasets (Behrens *et al.*, 2010). RF requires two tuning hyper-parameters, *i.e.* the number of randomly sampled variables at each split (m_{try}) and the number of regression trees (n_{tree}) grown in the forest. For each depth, the tuning of these parameters was done by iterating m_{try} values (*i.e.* the number of covariates) from 1 to 16 and n_{tree} values from 500 to 3000 with an increment of 100 (Hengl *et al.*, 2015). The “randomForest” and “caret” packages (Liaw *et al.*, 2002) of the R software were used for this goal.

Table 1. Pierson correlation analysis between soil physio-chemical properties and SAS indices

Soil properties	pH	EC	CCE	SOM	Clay	MWD	GMW	WSA
pH	1							
EC	0.16	1						
CCE	0.39**	0.16	1					
SOM	-0.59**	-0.06	-0.30**	1				
Clay	-0.38**	-0.06	-0.19*	0.77**	1			
MWD	-0.53**	-0.08	-0.29**	0.95**	0.82**	1		
GMW	-0.52**	-0.10	-0.29**	0.92**	0.84**	0.98**	1	
WSA	-0.49**	-0.02	-0.27**	0.87**	0.71**	0.85**	0.83**	1

*Correlation coefficients are significant at the $p < 0.05$ probability level. **Correlation coefficients are significant at the $p < 0.01$ probability level.

Table 2. Environmental covariates for soil aggregate stability indices prediction

Environmental covariates	Source	Covariates	Definition
Topographic attributes	Digital elevation model	DEM	Digital Elevation Model
		AnalyHil	Analytical Hillshading
		Convergence	Convergence index
		TWI	Topographic wetness index
Remote sensing attributes	Landsat-8 images	CNR	Carbonate Normalized Ratio (Red-Green/Red+Green)
		DVI	Difference Vegetation index
		INR	Iron Normalized Ratio (Red-SWIR2/Red-SWIR2)
	Sentinel-2 images	TSAVI	Transformed soil adjusted vegetation index $s(\rho\text{NIR}-s \times \rho\text{Red}-a)/a \times \rho\text{NIR} + \rho\text{Red} - a \times s + X \times (1 + s \times s)$
		CNR.s	Carbonate Normalized Ratio (Red-Green/Red+Green)
		CTVI	Corrected transformed vegetation index
		Brightness	Brightness index $((\text{Red})^2 + (\text{NIR})^2)^{0.5}$
		Wetness	Wetness index
		NDVI	Normalized difference vegetation index $(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
		Bands	Single Bands of Landsat-8 OLI(Banda3,11)
Landuse map	Land use	Land use	Land use map
Geology map	Geology	Geo	Geology map (1:100 000)
Soil variables	Lab analysis	pH, EC, CCE, SOM, Clay	pH, electrical conductivity, calcium carbonate equivalent, soil organic matter

2.4.2. Nearest Neighbor (k-NN)

The k-Nearest Neighbors (kNN) regression model is a versatile algorithm suitable for various tasks, including DSM. It is a non-parametric technique utilized for regression analysis. In this method, the prediction of a new data point is made by identifying the k nearest data points in the training set based on a distance metric, as documented by Nemes *et al.* (2006) and Zolfaghari *et al.* (2013). The prediction for the new point is then determined by averaging the values of the k nearest neighbors. We applied this method to model SAS indices, such as *MWD*, *GMD*, and *WSA*.

2.4.3. Support vector regression (SVR)

In the 1990s, support vector machine classification was proposed as a new approach in the learning machine theory by Chen *et al.* (2023). This method is a non-parametric supervised statistical method that works based on the assumption that there is no information about how the data set is distributed. The support vector regression (SVR) model is an adaptation of the SVM model for regression

tasks, commonly used in DSM as a ML model. In contrast to traditional regression models, SVR defines an error range, where predicted values within this range are deemed accurate. The size of the error interval determines the structure of the regression model (Zhao *et al.*, 2023). In this research, we utilize the "svmLinear2" function to establish correlations between SAS and various soil and environmental covariates in the context of DSM.

2.4.4. Artificial neural network (ANN)

The multi-layer perceptron algorithm with three layers was used to estimate the target data (test data). This type of algorithm consists of an input layer in order to apply problem inputs to train the network with data of auxiliary variables, a hidden layer with seven neurons and a sigmoid operator function, and an output layer with a linear operator function and will be able to make an estimate of SAS of the tested soils according to the input database (Behrens *et al.*, 2005). The neural network model was implemented using different numbers of environmental covariates.

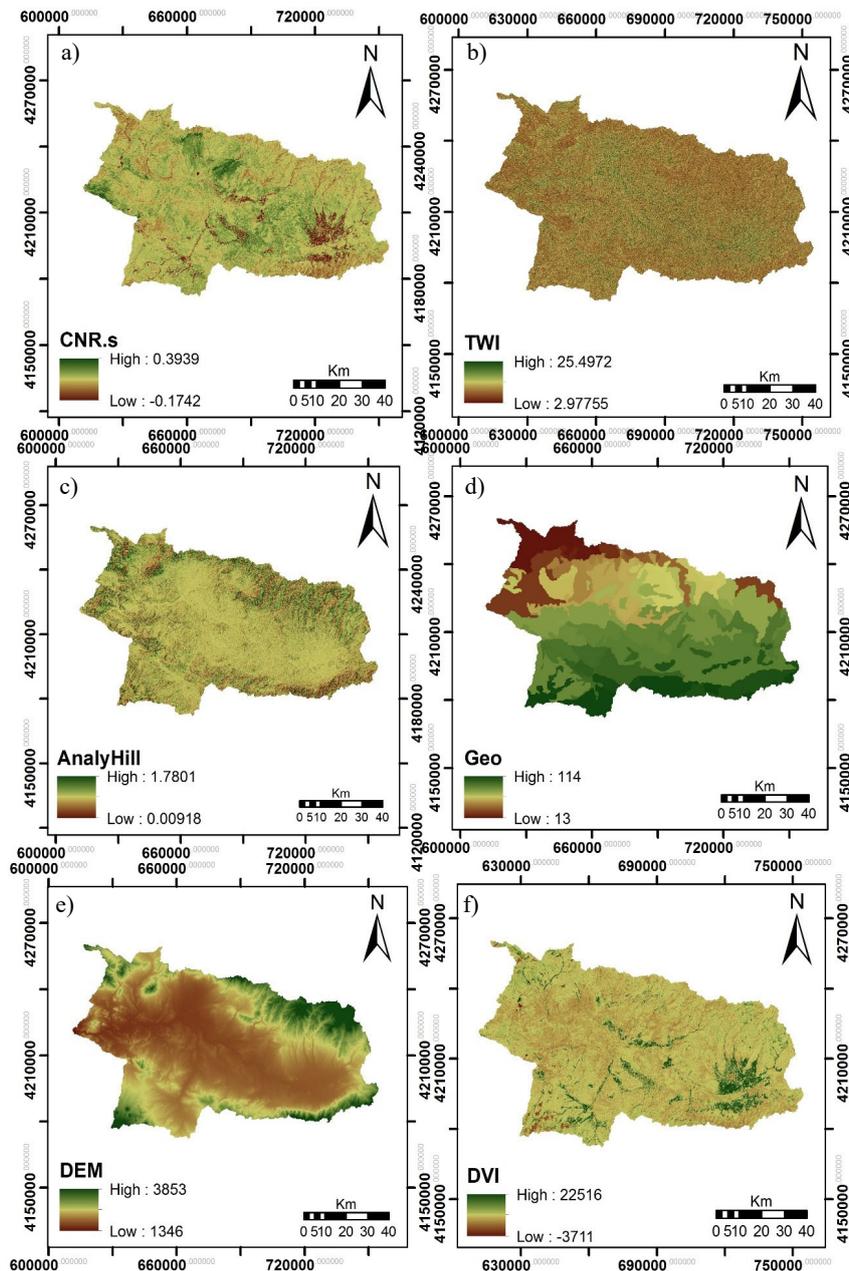


Fig. 2. Six of the most influential environmental covariates derived from remote sensing and topographic indices, a) carbonate Normalized ratio (Sentinei2), b) topographic wetness index, c) analytical Hillshading, d) geology, e) digital elevation model, f) difference vegetation index.

Table 3. Descriptive statistics of the scenarios of this study

Scenario	Abbreviate	Covariates
1	S ₁	Topographic attributes
2	S ₂	Topographic attributes+RS data
3	S ₃	Topographic attributes+RS data+ Geology+Land use/ cover
4	S ₄	Topographic attributes+RS data+Geology+Land use/ cover+ soil variables

2.4.5. Cubist (CB)

Cubist is a piecewise linear tree model that uses a recursive partitioning of the predictor covariates space (Ma *et al.*, 2017). It uses a divide-and-conquer strategy and seeks to minimize the intra subset variation at each node. CB models take the form: if (conditions) then (linear model) approach. If the predictor variables associated with an observation satisfy a set of conditions, the linear model is used to predict the response. The advantage of the condition set in each rule is that they enable interactions to be handled automatically by allowing different linear models to capture the local linearity in various parts of the predictor variables space. This can often lead to smaller trees and better prediction accuracy than regression trees (Ma *et al.*, 2017). CB requires tuning two parameters, *i.e.* the optimal number of committees and neighbors. The “caret” package of the R software was used for this goal.

2.5. Evaluation and assessment of models

For the aim of external validation of the various ML algorithms, the data set was randomly split into two parts, with 80% of them used for calibration and 20% used for validation. For tuning the RF, k-NN, SVR, ANN, and CB parameters, a 10-fold cross-validation was repeated 10 times (Heung *et al.*, 2016) using just for the calibration data set. Common statistical indices, *i.e.* coefficient of determination (R^2), Lin's concordance correlation coefficient (CCC), root mean square error ($RMSE$), and the percentage of normalized root mean squared error ($nRMSE$), were used to evaluate each model based on the validation data set using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}, \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}, \quad (6)$$

$$CCC = \frac{2r\sigma_{P_i}\sigma_{O_i}}{\sigma_{P_i}^2 + \sigma_{O_i}^2 + (\bar{P}_i - \bar{O}_i)^2}, \quad (7)$$

$$nRMSE = \frac{RMSE}{\bar{O}} 100, \quad (8)$$

where: P_i and O_i are the predicted and observed values, \bar{O} is the average of the observed values over the n measurements, r is the correlation coefficient between the predicted and observed values, and σ_P^2 , σ_O^2 is the variance of the predicted and observed values.

3. RESULTS AND DISCUSSION

3.1. Summary statistics and correlation analysis

The descriptive statistics of SAS indices and soil physicochemical properties are shown in Table 4. As can be seen, MWD varied from 0.13 to 2.92 mm with a mean of 1.08 mm. GMD varied from 0.11 to 2.16 mm with a mean of 0.86 mm, and WSA varied from 30.02 to 92.37% with a mean of 50.68%. Khazaei *et al.* (2008) reported a mean of 1.60 mm for MWD in some soils of Hamadan province, Iran. The arid and semi-arid climates of these two provinces do not provide suitable conditions for soil aggregate formation and stabilization. According to the Le Bissonnais (2016) classification, the aggregates in the studied soils have medium stability.

The lowest and highest coefficients of variance (CV) were obtained for pH and EC, and the low variability of pH could be due to the logarithmic nature of pH. SOC, Clay, CCE, MWD , and GMD showed high variability ($CV \geq 35\%$), while WSA had moderate variability ($15\% \leq CV < 35\%$) according to the Wilding and Drees (1978) classification. High variation of aggregate stability indices is due to different land management as well as variability in intrinsic soil properties, such as clay content and clay types (Havaee *et al.*, 2015). Zeraatpisheh *et al.* (2021) reported the highest CV for SOM and MWD and the lowest one for GMD and WSA in northern Iran. Similarly, several studies reported

Table 4. Descriptive statistics of soil aggregate stability indices in all studied soils (N = 100)

Variable	Unit	Mean	Max	Min	Median	Variance	StDev	CV (%)
pH	–	7.45	8.08	5.77	7.54	0.16	0.404	5.42
EC	(dS m ⁻¹)	0.536	3.74	0.06	0.23	0.59	0.77	143.65
CCE		9.57	48	0.2	7.5	78.26	8.84	92.37
SOM	(%)	1.13	4.5	0.1	0.8	0.83	0.911	80.61
Clay		22.44	40	1.6	20	130.6	11.42	50.89
MWD	(mm)	1.08	2.92	0.13	0.82	0.55	0.74	68.51
GMD		0.86	2.16	0.11	0.75	0.27	0.52	60.46
WSA	(%)	50.68	92.37	30.02	47.75	209.55	14.47	28.55

Min – minimum, Max – Maximum, StDev – standard deviation, CV – coefficient of variance, EC – electrical conductivity, CCE – calcium carbonate equivalent, SOM – soil organic matter, MWD – mean weight diameter of water-stable aggregates, GMD – geometric mean diameter of water-stable aggregates WSA – percentage of water stable aggregates.

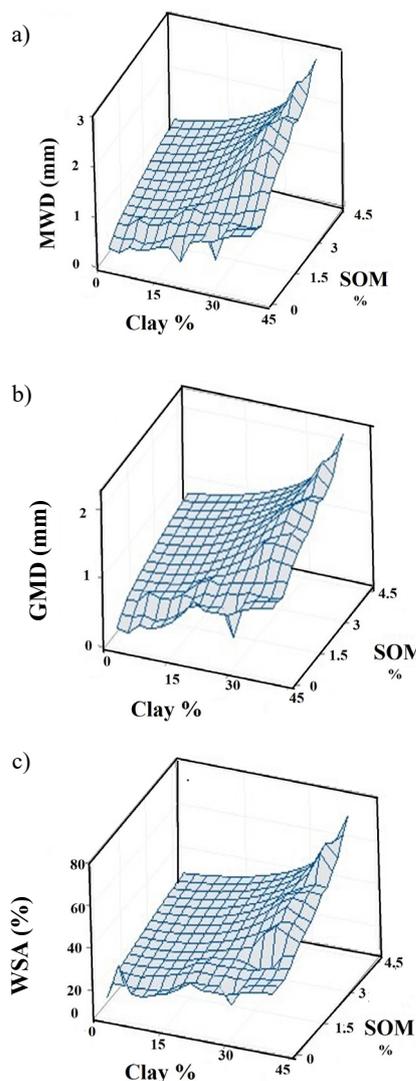


Fig. 3. a) Partial dependence plots between *MWD* with clay content and SOM, b) *GMD* with clay content and SOM, and c) *WSA* with clay content and SOM in the given watershed.

that land use changes and the conversion of natural ecosystems to cultivated lands have drastically altered aggregate stability and associated soil properties (Bakhshandeh *et al.*, 2019; Zeraatpisheh *et al.*, 2020; Ayoubi *et al.*, 2020).

The partial dependence plots between the SAS with soil properties, including the content of clay and SOM (Fig. 3), indicates that with the increase in the amount of SOM and, to some extent, clay, the percentage of the SAS increases, indicating the key role of SOM and clay (Table 1) in the SAS (Hati *et al.*, 2007).

3.2. Modeling performance

The final modeling process was assessed using four statistical indicators: coefficient of determination (R^2), Lin's concordance correlation coefficient (*CCC*), root mean square error (*RMSE*), and normalized root mean square

error (*nRMSE*). The performance of five ML models (RF, CB, SNM, ANN, and K-NN) in the four compiled scenarios for predicting SAS indices is presented in Tables 5-8. The results of modeling showed that the RF model for S_4 exhibited the highest accuracy in predicting the SAS indices, as evidenced for *MWD* by the lowest *RMSE* (0.33), *nRMSE* (31.42 %), and the highest R^2 (0.86) and *CCC* (0.87) in the validation dataset, while the following values were obtained in the *GMD* prediction: *RMSE* (0.26), *nRMSE* (30.59%), R^2 (0.80), and *CCC* (0.84). Furthermore, for the *WSA* prediction, the lowest *RMSE* (10.55) and *nRMSE* (20.75%) as well as R^2 (0.54) and *CCC* (0.68) were obtained (Table 8). The RF model turned out to be a successful model in the prediction of SAS indices because of several drivers, such as the size of the dataset, the scale of variation, and the relationships between dependent and independent covariates (da Silva Chagas *et al.*, 2016). In the analysis of the other scenarios, the results showed that the CB model is highly accurate for the SAS in scenario 1 (S_1) (Table 5), while in scenario 2 (S_2), the RF, SVR, and CB models had the highest accuracy for the *MWD*, *GMD*, and *WSA* indicators, respectively (Table 6). Also, in scenario 3 (S_3), the SVR models for predicting *MWD* and *WSA* and the CB model for estimating *GMD* had the best efficiency and accuracy (Table 7). Finally, in scenario 4 (S_4), the RF model had the highest accuracy for SAS evaluation due to its high efficiency. RF is insensitive to noise or over-training, which shows its ability in dealing with unbalanced data (Boateng *et al.*, 2020).

Mousavi *et al.* (2022) and Rahmani *et al.* (2022) confirmed that the RF algorithm had high accuracy and low error for predicting SOC and topsoil thickness in Qazvin plain, Iran. Shi *et al.* (2020) reported an R^2 of 0.50 when aggregate stability was estimated using high-resolution Airborne Prism Experiment hyperspectral images in croplands. Bouslihim *et al.* (2021b) found that the RF model resulted in the lowest *RMSE* (<0.15) and higher R^2 (>0.92), compared to multiple linear regression for predicting SAS indices. Zeraatpisheh *et al.* (2021) also showed that the RF model with higher R^2 (0.75, 0.74, and 0.58) and lower *nRMS* (24.28, 12.72, and 10.40) performed better for predicting SAS indices, while the K-NN and SVR models showed a weaker performance.

The CB model also "as the second best model" demonstrated the most accurate predictions for the *MWD*, *GMD*, and *WSA* indices for S_4 , as indicated by the lowest *RMSE* (0.38, 0.27, and 10.04, respectively) and *nRMSE* (36.19, 31.76, and 19.75 respectively) as well as the highest R^2 (0.77, 0.78, 0.51) and *CCC* (0.83, 0.84, and 0.62,) respectively (Table 8). The priority of models regarding their performance for predicting *MWD*, *GMD*, and *WSA* were as follows: RF > CB > SVR > ANN > K-NN. In contrast, the K-NN model exhibited the weakest prediction accuracy for *MWD*, *GMD*, and *WSA*, with the lowest R^2 and *CCC* values. According to Malone *et al.* (2009), the RF, CB, and SVM

Table 5. Validation criteria for prediction of aggregate stability indices of all models for scenario 1 (S_1)

ML model	R ²	CCC	RMSE	nRMSE(%)
<i>MWD</i>				
RF	0.1	0.25	0.67	63.80
CB	0.12	0.33	0.78	74.28
SVM	0.10	0.21	0.81	77.14
ANN	0.15	-4.58	0.26	74.28
K-NN	0.08	0.17	0.64	60.95
<i>GMD</i>				
RF	0.11	0.27	0.51	60
CB	0.14	0.35	0.54	63.52
SVM	0.11	0.25	0.87	94.56
ANN	0.16	-4.93	0.24	66.66
K-NN	0.12	0.21	0.46	54.11
<i>WSA</i>				
RF	0.08	0.19	11.66	22.94
CB	0.10	0.24	11.91	23.43
SVM	0.10	0.22	11.48	22.58
ANN	0.06	-4.19	0.23	69.69
K-NN	0.06	0.14	13.58	26.72

RF – random forest, CB – Cubist, SVM – support vector machine, ANN – artificial neural network, K-NN – Nearest Neighbor.

models performed moderately in predicting SAS, while ANN along with K-NN demonstrated weaker results. Ye *et al.* (2018) reported a R² value of 0.40, comparable with our results, and a higher RMSE of 0.53 for MWD of a small catchment on the Loess plateau in China. In accordance with our findings, Khaledian and Miler (2020) in a review regarding selecting appropriate machine learning methods for DSM, found that, in almost all studies in spatial modeling of soil properties, the CB model had approximately similar performance to that of RF in terms of R² and RMSE.

Our results showed that nRMSE were less than 42% for the best estimation (Table 8). The nRMSE close to 40% means acceptable prediction accuracy, while the values higher than 71% show unacceptable prediction accuracy (Hengl *et al.*, 2004; Kamamia *et al.*, 2021). Besalatpour *et al.* (2013) used four different models, including the adaptive network-based fuzzy inference system (ANFIS), the generalized linear model (GLM), ANNs, and multi linear regression (MLR), to predict the MWD in a highly mountainous watershed in Iran and found lower accuracy criteria than in the current study. They showed that the R² value in the MLR model ranged from 0.07 to 0.18 for three different sets (soil data, vegetation and topographic data, and the combination of the three covariates). In the same way,

Table 6. Validation criteria for prediction of aggregate stability indices of all models for scenario 2 (S_2)

ML model	R ²	CCC	RMSE	nRMSE(%)
<i>MWD</i>				
RF	0.60	0.64	0.46	42.60
CB	0.36	0.33	0.73	67.59
SVM	0.45	0.50	0.60	55.55
ANN	0.27	0.46	0.26	78.78
K-NN	0.26	0.47	0.46	42.59
<i>GMD</i>				
RF	0.36	0.37	0.46	54.11
CB	0.40	0.33	0.50	58.82
SVM	0.41	0.51	0.42	49.41
ANN	0.40	0.48	0.18	50
K-NN	0.32	0.47	0.37	43.52
<i>WSA</i>				
RF	0.40	0.58	8.10	15.93
CB	0.45	0.64	8.83	17.40
SVM	0.44	0.47	12.52	24.63
ANN	0.32	0.47	0.18	54.54
K-NN	0.23	0.25	14.34	28.21

Explanations as in Table 5.

Asadi and Bagheri (2010) attempted to predict SAS with the ANN and MLR models in Iran. The obtained R² values for the MLR model ranged from 0.15 to 0.39, which is lower than the results obtained in the current study. This confirms that the relationships between SAS indices and environmental covariates are not exactly linear, and some of the relationships might be non-linear, which is explained by non-linear models, such as RF and ANN.

Among the five applied MLMs (RF, CB, ANN, K-NN, and SVR), our results showed that the K-NN and ANN models had the weakest performance for all the scenarios, compared to the other models for predicting MWD, GMD, and WSA, respectively. Several researchers (*i.e.* Mansuy *et al.*, 2014; Taghizadeh-Mehrjardi *et al.*, 2016; Campbell *et al.*, 2018; Gunarathna *et al.*, 2019; Szabó *et al.*, 2019) using MLM in soil mapping revealed that in all sample sizes, the K-NN model had lower performance than the RF and CB models. Boateng *et al.* (2020) reported that a part of this weakness for the KNN model might be attributed to the high difficulty of setting the ideal value of K for the KNN classifier, although this algorithm was easy to implement and understand. In the comparison of the KNN, SVR and RF models, Boateng *et al.* (2020) stated that, due to the numerous types of NN architectures to choose from and

Table 7. Validation criteria for prediction of aggregate stability indices of all models for scenario 3 (S_3)

ML model	R ²	CCC	RMSE	nRMSE(%)
<i>MWD</i>				
RF	0.40	0.44	0.63	60
CB	0.38	0.45	0.62	59.04
SVM	0.44	0.41	0.70	66
ANN	0.25	0.39	0.23	65.71
K-NN	0.15	0.25	0.70	66.66
<i>GMD</i>				
RF	0.42	0.53	0.38	44.70
CB	0.50	0.62	0.35	40.69
SVM	0.46	0.50	0.44	51.76
ANN	0.21	0.37	0.21	58.33
K-NN	0.14	0.25	0.47	55.29
<i>WSA</i>				
RF	0.41	0.52	11.66	22.94
CB	0.32	0.54	11.95	20.68
SVM	0.38	0.51	13.90	27.35
ANN	0.2	0.33	14.08	42.66
K-NN	0.19	0.22	15.36	30.22

Explanations as in Table 5.

the high number of algorithms used for training, most researchers recommend SVR and RF as easier and wieldy used methods which repeatedly achieve results with high accuracies and are often faster to implement. Khaledian and Miler (2020) compared various ML models in digital soil mapping and claimed that various factors, such as sample size, covariate numbers, and learning time, affected the performance of the models. They accentuated that ANN models were more powerful when the datasets were large, and for a dataset less than 100, RF, SVR and CB were more efficient.

3.3. Relative importance analysis of covariates

The relative importance (RI) analysis for the best performing model and scenario in predicting SAS indices is visually depicted in Fig. 4. The findings revealed that soil properties, RS data, and topographic data significantly influenced the variability of the SAS indices. According to Fig. 3, SOM, clay, CNRs, and TWI were recognized as four top covariates for predicting SAS indices by explaining 59, 58, and 60% of variability in *MWD*, *GMD*, and *WSA*, respectively. SOM and clay are two important properties that affect the stability of soil aggregates, because a large portion of SOM contains particulate organic mate-

Table 8. Validation criteria for prediction of aggregate stability indices of all models for scenario 4 (S_4)

ML model	R ²	CCC	RMSE	nRMSE(%)
<i>MWD</i>				
RF	0.86	0.87	0.33	31.42
CB	0.77	0.83	0.38	36.19
SVM	0.73	0.8	0.39	37.14
ANN	0.11	0.19	0.34	97.14
K-NN	0.11	0.16	0.80	76.19
<i>GMD</i>				
RF	0.8	0.84	0.26	30.59
CB	0.78	0.84	0.27	31.76
SVM	0.74	0.82	0.30	35.29
ANN	0.15	0.22	0.30	83.33
K-NN	0.16	0.20	0.55	64.70
<i>WSA</i>				
RF	0.54	0.68	10.55	20.75
CB	0.51	0.62	10.04	19.75
SVM	0.53	0.65	10.22	20.11
ANN	0.08	0.17	0.28	84.84
K-NN	0.02	0.11	15.61	30.71

Explanations as in Table 5.

rials (Soinnie *et al.*, 2016), and they are substantially influenced by land use and land management. Research conducted by Liu *et al.* (2023) highlighted the crucial role of SOM in controlling soil aggregation. Djukic *et al.* (2010) found that changes in precipitation, temperature, vegetation types, and SOM decomposition greatly impacted SOM pools with elevation. The bonding between clay particles and organic matter plays a vital role in SAS, as demonstrated by Hati *et al.* (2007). Moreover, an increase in clay and organic matter content in soil enhances resistance in SA (Annabi *et al.*, 2017). Khazaei *et al.* (2008) also emphasized the significant influence of SOM, clay, and calcium carbonate on SAS. Overall, organic matter plays a key role in enhancing soil structural stability through such mechanisms as increasing hydrophobicity and particle adhesion. Several studies (Annabi *et al.*, 2017; Chaney and Swift, 1984; Chenu *et al.*, 2000; Kavdir *et al.*, 2004) confirm the substantial impact of SOM and clay on SAS indices. Among the topographic attributes, topographic wetness index (TWI) had a high contribution to predicting SAS indices, as shown in Fig. 4. Cantón *et al.* (2009), Tang *et al.* (2010), and Nsabimana *et al.* (2020) revealed that TWI, elevation, and slope played a significant role in the spatial distribution of SAS. Additionally, topography

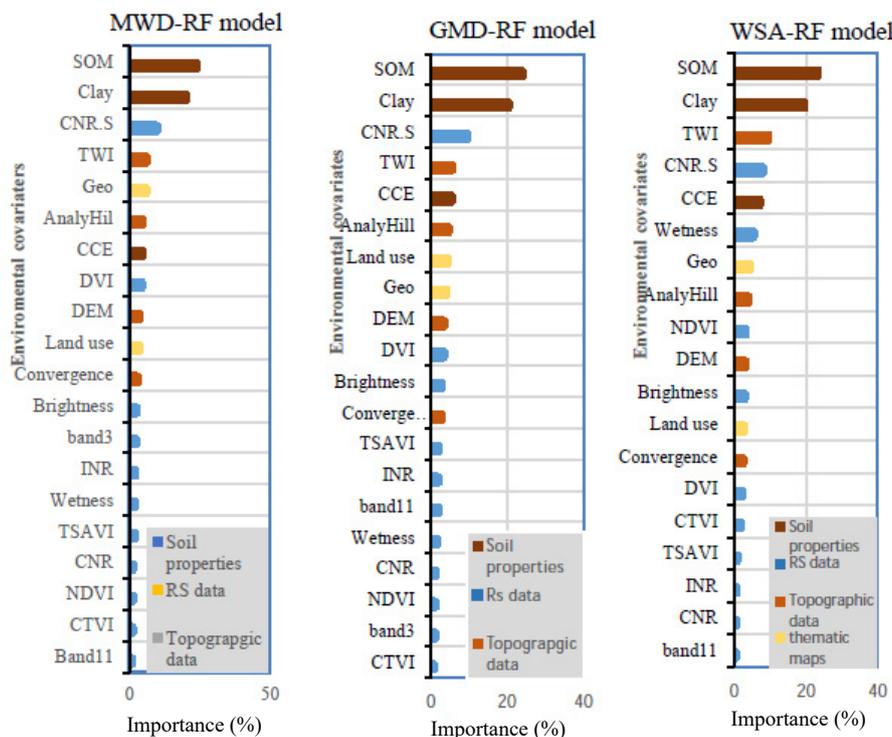


Fig. 4. Relative importance of environmental covariates (%) for predicting SAS indices for the best prediction model in scenario 4 (S_4).

has vital impacts on other soil properties and vegetation distribution, which inevitably influence SAS. For instance, Adhikari *et al.* (2018) emphasized the critical role of parameters derived from DEM in the spatial distribution of soil properties that regulate SA. Moreover, the results of Wang *et al.* (2023) demonstrated significant differences in soil functions among different land use types and slope positions, highlighting the impact of slope position on SA. Furthermore, the TWI, as a physical indicator of the influence of regional topography on runoff flow and water storage, has been shown to be the most important variable for predicting SOC (Taghizadeh-Mehrjardi *et al.*, 2016; Tajik *et al.*, 2020). The Carbonate Normalized Ratio derived from Sentinel satellite (CNRs) was found to be a proxy for RS data in predicting SAS indices (Fig. 3). Additionally, RS data and its derivative covariates, along with topographic attributes, have been identified as significant predictors for predicting SAS (Zeraatpisheh *et al.*, 2020; Jones *et al.*, 2021). RS spectral data reflects land surface conditions and is relevant for exploring soil properties, such as SOC, soil moisture, and soil texture (Browning and Duniway, 2011). Several scholars have confirmed the influential role of RS data in predicting soil properties and SA aggregation (Minasny *et al.*, 2013; Zeraatpisheh *et al.*, 2019; Tajik *et al.*, 2020; Matinfar *et al.*, 2021; Mahjenabadi *et al.*, 2022). The CNR.s index was reported as the third most important predictor for *MWD* and *GMD* and the fourth for *WSA*, indicating the crucial contribution of carbonate-rich parent materials to SAS indices viability. This

index is the consequence of soil carbonate effects on soil aggregation and ultimately on *MWD*, *GMD*, and *WSA*. Following the CNRs among the RS indices, the brightness index, DVI, NDVI, and TSVI with different weights had significant contributions to explaining the variability of *MWD*, *WSA*, and *GMD*. Vegetation indices are regarded as proxies for native vegetation cover that directly or indirectly influences SAS indices (Jones *et al.*, 2021). Overall, the results of the covariate importance analysis showed that soil properties, topographic data, and RS data derived from Sentinel 2 in the last scenario (S_4) had a high contribution to explaining the variability of SAS and their estimation at the watershed scale. In this regard, Khosravani *et al.* (2023) investigated the prediction of SOC and soil properties in two scenarios and showed that soil variables along with environmental covariates improved the accuracy of MLMs, compared to the scenario without soil properties. Similarly, Zeraatpisheh *et al.* (2021) demonstrated that soil and RS covariates were recognized as the most important driving factors of the SAS. Also, Kamamia *et al.* (2021) found that soil covariates, remote sensing, and DEM variables were recognized as the most effective factors in explaining *MWD* in Kenya.

3.4. Spatial prediction of soil aggregate stability with uncertainty estimates

Since the RF model demonstrated high accuracy in predicting SAS indices, we utilized spatial prediction and uncertainty maps to illustrate the predicted SAS indices using this model and associated uncertainty maps for three

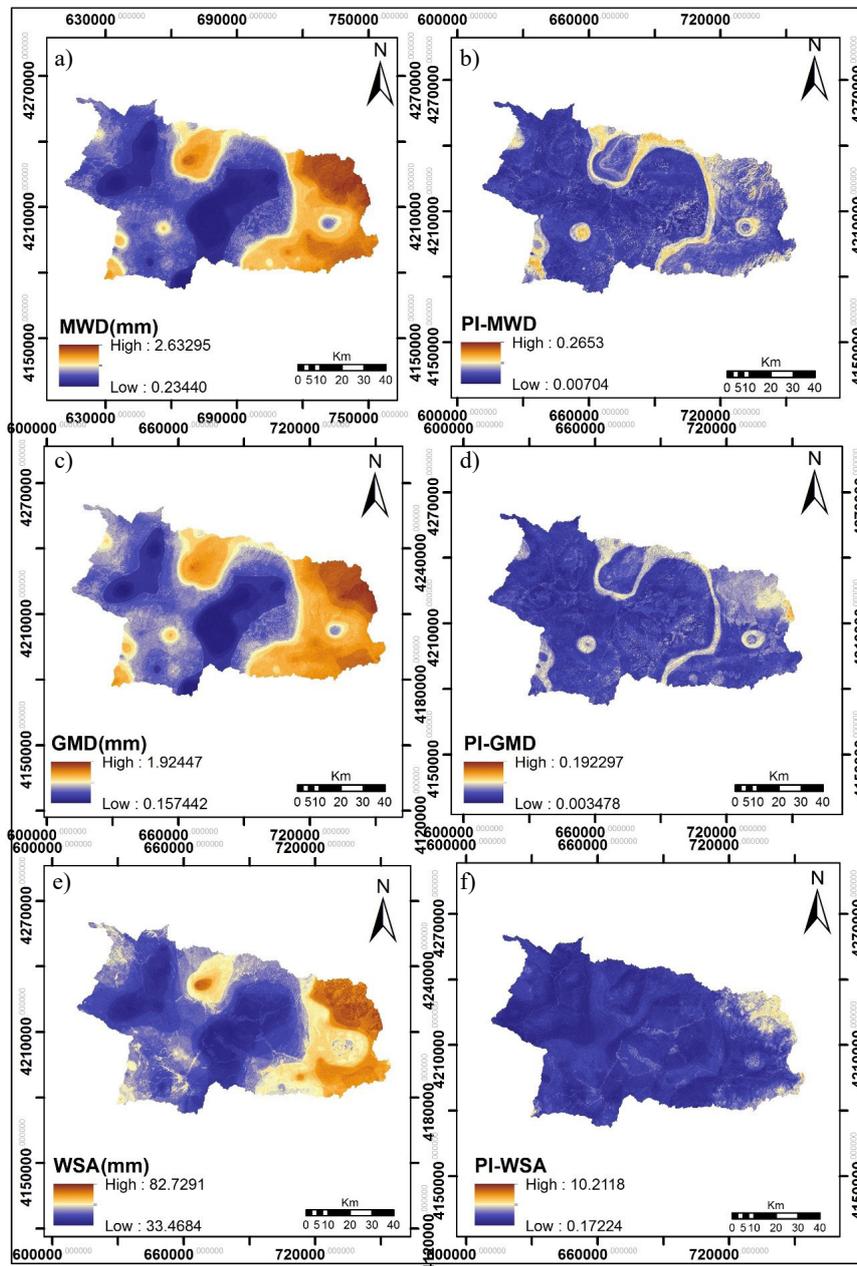


Fig. 5. Spatial prediction of different SAS indices for the best prediction model in scenario 4 (S_4): a) mean weight diameter (MWD) predicted by the random forest (RF) model, b) uncertainty map (PI- MWD), c) the geometric mean diameter (GMD) predicted by the random forest (RF) model, d) Uncertainty map (PI- GMD), e) water-stable aggregates (WSA) predicted by the random forest (RF) model, f) uncertainty map (PI- WSA).

studied indices. As shown in Fig. 5a, the spatial predictions of MWD in the northeastern, northwestern, some southeastern, and southwestern areas of the study region, characterized by higher altitude and pasture land with good vegetation cover, exhibit the highest MWD contents. Conversely, the central and western regions, dominated by agricultural land use, particularly long-term cultivation, and pastures with poor vegetation, show the lowest MWD values. Several scholars reported that intensive cultivation and tillage practices led to the destruction of macro-aggregates

and the enhancement of micro-aggregates, and subsequently lowering MWD (Ayoubi *et al.*, 2012; Weidhuner *et al.*, 2021; Ferreira *et al.*, 2023).

To assess the model's uncertainty in predicting MWD , we calculated the ratio of measured soil values within the 90% prediction interval to the average prediction interval (Shrestha and Solomatine, 2006). Ideally, 90% of observations should fall within the defined prediction interval with a 90% confidence level, and the interval should be as narrow as possible. This uncertainty analysis partly confirms

the model's ability to predict *MWD*, as shown in Fig. 5b. Additionally, Fig. 5c illustrates the spatial prediction of *GMD* using the RF model, indicating higher values in the eastern and northwestern parts, with lower values in the central and western areas due to agricultural practices. Figure 5d demonstrates the high accuracy of the model's prediction uncertainty map for this index. Furthermore, Fig. 5e displays the *WSA* prediction map, the highest values are noted in the areas under pasture land use and the lowest values are recorded in the agricultural land and the barren areas, respectively. The destruction of aggregates and their transformation into fine aggregates in the arable land reduces the SAS due to agricultural operations. On the other hand, the wetting of the soil aggregates in these land use systems causes air to be trapped between the soil aggregates and increases the dispersion of clay particles, and as a result, *WSA* decreases (Gholoubi *et al.*, 2019). However, in pasture land use, there are more stable aggregates due to the greater content of organic carbon and clay, structural porosity, and a lack of cultivation. The uncertainty map of the RF model for the *WSA* index is shown in Fig. 5f. The highest values of SAS in the pasture soils are in accordance with high clay and higher organic matter contents in these soils that enhance the aggregation, while all the three indices had the lowest values in the agricultural and barren soils with lower clay and organic matter levels.

Overall, the SAS assessment is a function of the amount of carbon in the soil and that, in turn, is a function of the input of plant residues and their loss from the soil (Fuentes *et al.*, 2004). In pasture soil, due to a lack of cultivation and higher content of plant residues, there is a balance between the decomposition of SOM and the accumulation of plant residues. However, this balance is disturbed in arable lands because of the harvest of plant biomass and the lack of a return of plant residues to the soil (Wang *et al.*, 2014). Also, in arable lands, tillage operations cause the mixing of the lower layers of the soil with a lower percentage of organic carbon with the surface soil containing more organic carbon, and as a result, they cause a decrease in soil organic carbon, which plays a great role in flocculating soil particles and then stabilizing the soil aggregate (Tejada and Gonzalez, 2008). Many studies showed that long-term cultivation and agricultural practices could decrease the amount of SOM (Solomon *et al.*, 2002; Tajik *et al.*, 2019a, 2019b) and, consequently, decline soil aggregate stability indices. Similarly, Celik (2005) found 61% decreases in *MWD* for the surface layer, and Caravaca *et al.* (2004) reported 40% decreases in *WSA* in cultivated soil, compared with the natural forest soil.

Generally, the results showed that, in general, MLMs have a high ability to predict the SAS indices at the watershed scale. They accurately predicted a range of the SAS indices using environmental variables in combination with easily measured soil properties, such as clay and SOM. It seems that using satellite images with higher spectral

and spatial resolution as well as DEM with higher spatial resolution enhance the spatial prediction of SAS at the watershed scale.

4. CONCLUSIONS

This study aimed to employ five multi linear regressions to predict the spatial distribution of soil aggregate stability indices of the surface soil (0-15 cm depth) under four scenarios combining various auxiliary variables, such as topographic attributes, remote sensing and thematic maps, with easily available soil properties in northwestern Iran.

The results of the validation showed that the random forest model in scenario 4 (S_4) exhibited very well predictions for mean weight diameter and geometric mean diameter. However, $R^2=0.54$ for water-stable aggregates is definitely not a good achievement. These results confirmed the potential of the random forest model in digital soil mapping studies, and the combinations of a series of auxiliary variables that cover a high range of soil forming factors have higher efficacy for predicting complex variables such as soil aggregate stability. The uncertainty analysis of the prediction maps further confirmed the RF model's accuracy and reliability in predicting these indicators.

Among the selected covariates, soil properties, remote sensing data, and topographic attributes emerged as the most influential factors in predicting soil aggregate stability indicators in unobserved areas. Specifically, soil organic matter and clay content demonstrated a significant influence in predicting soil aggregate stability. It is essential to acknowledge that, while these covariates captured a substantial portion of spatial variation in soil aggregate stability, they may not encompass all factors influencing soil aggregate stability. Furthermore, the soil aggregate stability maps highlighted the significant role of land use in soil aggregate stability, particularly through the influence of organic carbon. The higher content of soil organic matter and clay in pasture soils, compared to arable land, underscored the impact of land use on soil organic matter and soil aggregation.

In a broader context, these findings emphasize the potential of mapping methods as a viable approach for modeling soil aggregate stability indices, thereby facilitating improved management and erosion control in the study area. This underscores the broader applicability of such modeling approaches in enhancing soil quality and sustainability across diverse geographical regions.

5. ACKNOWLEDGMENT

The authors greatly acknowledge the Isfahan University of Technology for the financial support of this research.

Conflicts of Interest: The Authors do not declare any conflict of interest.

6. REFERENCES

- Adhikari, K., Owens, P.R., Ashworth, A.J., Sauer, T.J., Libohova, Z., Richter, J.L. and Miller, D.M., 2018. Topographic controls on soil nutrient variations in a silvopasture system. *Agrosys. Geosci. Environ.* 1(1), 1-15. <https://doi.org/10.2134/age2018.04.0008>
- Annabi, M., Raclot, D., Bahri, H., Bailly, J.S., Gomez, C., Le Bissonnais, Y., 2017. Spatial variability of soil aggregate stability at the scale of an agricultural region in Tunisia. *Catena* 153, 157-167. <https://doi:10.1016/j.catena.2017.02.010>
- Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A.C., Morgan, C.L., Roudier, P., Poggio, L., Mulder, V.L., 2020. Impressions of digital soil maps: The good, the not so good, and making them ever better. *Geoderma Regional* 20, p.e00255. <https://doi.org/10.1016/j.geodrs.2020.e00255>
- Asadi, H., Bagheri, F., 2010. Comparison of regression pedotransfer functions and artificial neural networks for soil aggregate stability simulation. *World Appl. Sci. J.* 8(9), 1065-1072.
- Ayoubi, S., Karchegani, P.M., Mosaddeghi, M.R., Honarjoo, N., 2012. Soil aggregation and organic carbon as affected by topography and land use change in western Iran. *Soil Till. Res.* 121, 18-26. <https://doi.org/10.1016/j.still.2012.01.011>
- Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M.R., 2020. Soil organic carbon physical fractions and aggregate stability influenced by land use in humid region of northern Iran. *Int. Agrophys.* 34(3). <https://doi.org/10.31545/intagr/125620>
- Bakhshandeh, E., Hossieni, M., Zeraatpisheh, M., Francaviglia, R., 2019. Land use change effects on soil quality and biological fertility: a case study in northern Iran. *Eur. J. Soil Biol.* 95, 103119. <https://doi.org/10.1016/j.ejsobi.2019.103119>
- Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D. Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. *J. Plant Nutrition Soil Sci.* 168(1), 21-33. <https://doi.org/10.1002/jpln.200421414>
- Behrens, T., Zhu, A.X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma* 155(3-4), 175-185. <https://doi.org/10.1016/j.geoderma.2009.07.010>
- Besalatpour, A.A., Ayoubi, S., Hajabbasi, M.A., Mosaddeghi, M., Schulin, R., 2013. Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. *Catena* 111, 72-79. <https://doi.org/10.1016/j.catena.2013.07.001>
- Besalatpour, A.A., Ayoubi, S., Hajabbasi, M.A., Jazi, A.Y., Gharipour, A., 2014. Feature selection using parallel genetic algorithm for the prediction of geometric mean diameter of soil aggregates by machine learning methods. *Arid Land Res. Manag.* 28(4), 383-394. <https://doi:10.1080/15324982.2013.871599>
- Bhattacharyya, R., Rabbi, S.M., Zhang, Y., Young, I.M., Jones, A.R., Dennis, P.G., *et al.*, 2021. Soil organic carbon is significantly associated with the pore geometry, microbial diversity and enzyme activity of the macro-aggregates under different land uses. *Sci. Total Environ.* 778, 146286. <https://doi.org/10.1016/j.scitotenv.2021.146286>
- Boateng, E.Y., Otoo, J., Abaye, D.A., 2020. Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review. *J. Data Analysis Information Processing* 8(4), 341-357. <https://doi.org/10.4236/jdaip.2020.84020>
- Bouslihim, Y., Rochdi, A., Aboutayeb, R., El Amrani-Paaza, N., Miftah, A., Hssaini, L., 2021a. Soil aggregate stability mapping using remote sensing and GIS-based machine learning technique. *Frontiers Earth Sci.* 9, 748859. <https://doi.org/10.3389/feart.2021.748859>
- Bouslihim, Y., Rochdi, A., Paaza, N.E.A., 2021b. Machine learning approaches for the prediction of soil aggregate stability. *Heliyon* 7(3), <https://doi.org/10.1016/j.heliyon.2021.e06480>
- Breiman, L., 2001. Random forests. *Machine learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Browning, D.M., Duniway, M.C., 2011. Digital soil mapping in the absence of field training data: A case study using terrain attributes and semiautomated soil signature derivation to distinguish ecological potential. *Applied Environ. Soil Sci.* 2011(1), 421904. <https://doi.org/10.1155/2011/421904>
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 239, 68-83. <https://doi.org/10.1016/j.geoderma.2014.09.019>
- Campbell, P.M.D.M., Fernandes Filho, E.I., Francelino, M.R., Demattê, J.A.M., Pereira, M.G., Pinto, L.A.D.S.R., 2019. Digital soil mapping of soil properties in the “Mar de Morros” environment using spectral data. *Revista Brasileira de Ciência do Solo*, 42, p.e0170413. <http://dx.doi.org/10.1590/18069657rbcs20170413>
- Cantón, Y., Solé-Benet, A., Asensio, C., Chamizo, S., Puigdefàbregas, J., 2009. Aggregate stability in range sandy loam soils relationships with runoff and erosion. *Catena* 77(3), 192-199. <https://doi:10.1016/j.catena.2008.12.011>
- Caravaca, F., Lax, A., Albaladejo, J., 2004. Aggregate stability and carbon characteristics of particle-size fractions in cultivated and forested soils of semiarid Spain. *Soil Till. Res.* 78(1), 83-90. <https://doi.org/10.1016/j.still.2004.02.010>
- Celik, I., 2005. Land-use effects on organic matter and physical properties of soil in a southern Mediterranean highland of Turkey. *Soil Till. Res.* 83(2), 270-277. <https://doi.org/10.1016/j.still.2004.08.001>
- Chahal I., Eerd L.L., 2019. Quantifying soil quality in a horticultural-cover cropping system. *Geoderma* 352, 38-48. <https://doi.org/10.1016/j.geoderma.2019.05.039>
- Chaney, K., Swift, R.S., 1984. The influence of organic matter on aggregate stability in some British soils. *J. Soil Sci.* 35(2), 223-230. <https://doi.org/10.1111/j.1365-2389.1984.tb00278.x>
- Chaplot, V., Cooper, M., 2015. Soil aggregate stability to predict organic carbon outputs from soils. *Geoderma* 243, 205-213. <https://doi.org/10.1016/j.geoderma.2014.12.013>
- Chen, S., Arrouays, D., Mulder, V.L., Poggio, L., Minasny, B., Roudier, P., *et al.*, 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma* 409, 115567. <https://doi.org/10.1016/j.geoderma.2021.115567>
- Chen, D., Zhao, J., Qin, S., 2023. SVM strategy and analysis of a three-phase quasi-Z-source inverter with high voltage transmission ratio. *Sci. China Technol. Sciences* 66(10), 2996-3010. <https://doi.org/10.1007/s11431-022-2394-4>
- Chenu, C., Le Bissonnais, Y., Arrouays, D., 2000. Organic matter influence on clay wettability and soil aggregate stability. *Soil Sci. Soc. Am. J.* 64(4), 1479-1486. <https://doi.org/10.2136/sssaj2000.6441479x>
- da Silva Chagas, C., de Carvalho Junior, W., Bhering, S.B., Calderano Filho, B., 2016. Spatial prediction of soil surface

- texture in a semiarid region using random forest and multiple linear regressions. *Catena* 139, 232-240. <https://doi.org/10.1016/j.catena.2016.01.001>
- Deng, C., Teng, X., Peng, X., Zhang, B., 2014. Effects of simulated puddling intensity and pre-drying on shrinkage capacity of a paddy soil under long-term fertilization. *Soil Till. Res.* 140, 135-143. <https://doi.org/10.1016/j.still.2014.02.012>
- Dexter A.R., Richard G., Arrouays D., Czyż E.A., Jolivet C., Duval O., 2008. Complexed organic matter controls soil physical properties. *Geoderma* 144, 620-627. <https://doi.org/10.1016/j.geoderma.2008.01.022>
- Ding, J., Li, F., Yang, G., Chen, L., Zhang, B., Liu, L., *et al.*, 2016. The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores. *Global Change Biol.* 22(8), 2688-2701. <https://doi.org/10.1111/gcb.13257>
- Djukic, I., Zehetner, F., Tatzber, M., Gerzabek, M.H., 2010. Soil organic-matter stocks and characteristics along an Alpine elevation gradient. *J. Plant Nutr. Soil Sci.* 173(1), 30-38. <https://doi.org/10.1002/jpln.200900027>
- Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., Scholten, T., 2020. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sensing* 12(14), 2234. <https://doi.org/10.3390/rs12142234>
- Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. *Methods of soil analysis: Part 1. Physical and mineralogical methods*, 5, 383-411. <https://doi.org/10.2136/sssabookser5.1.2ed.c15>
- Geisseler, D., Scow, K.M., 2014. Long-term effects of mineral fertilizers on soil microorganisms – A review. *Soil Biol. Biochemistry*, 75, 54-63. <https://doi.org/10.1016/j.soilbio.2014.03.023>
- Gholoubi, A., Emami, H., Caldwell, T., 2019. Deforestation effects on soil aggregate stability quantified by the high energy moisture characteristic method. *Geoderma* 355, 113919. <https://doi.org/10.1016/j.geoderma.2019.113919>
- Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.K.N., 2019. Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. *Water* 11(9), 1940. <https://doi.org/10.3390/w11091940>
- Hanke, D., Dick, D.P., 2017. Aggregate stability in soil with humic and histic horizons in a toposequence under Araucaria Forest. *Revista Brasileira de Ciência do Solo* 41, <https://doi.org/10.1590/18069657rbcs20160369>
- Hati, K.M., Swarup, A., Dwivedi, A.K., Misra, A.K., Bandyopadhyay, K.K., 2007. Changes in soil physical properties and organic carbon status at the topsoil horizon of a vertisol of central India after 28 years of continuous cropping, fertilization and manuring. *Agric. Ecosys. Environ.*, 119(1-2), 127-134. <https://doi.org/10.1016/j.agee.2006.06.017>
- Havaee, S., Ayoubi, S., Mosaddeghi, M.R., Keller, T., 2014. Impacts of land use on soil organic matter and degree of compactness in calcareous soils of central Iran. *Soil Use Management* 30(1), 2-9. <https://doi.org/10.1111/sum.12092>
- Hengl, T., Heuvelink, G.B., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120(1-2), 75-93. <https://doi.org/10.1016/j.geoderma.2003.08.018>
- Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., *et al.*, 2015. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS one* 10(6), p.e0125814. <https://doi.org/10.1371/journal.pone.0125814>
- Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265, 62-77. <https://doi.org/10.1016/j.geoderma.2015.11.014>
- Farahani, E.M.A.M.I., Emami, H., Fotovat, A., Khorassani, R., 2019. Effect of different K: Na ratios in soil on dispersive charge, cation exchange and zeta potential. *Eur. J. Soil Sci.* 70(2), 311-320. <https://doi.org/10.1111/ejss.12735>
- Ferreira, T.R., Archilha, N.L., Cássaro, F.A., Pires, L.F., 2023. How can pore characteristics of soil aggregates from contrasting tillage systems affect their intrinsic permeability and hydraulic conductivity?. *Soil Till. Res.* 230, 105704. <https://doi.org/10.1016/j.still.2023.105704>
- Fuentes J.P., Flury M., Bezdicsek D., 2004. Hydraulic properties in a silt loam soil under natural prairie, conventional till, and no-till. *Soil Sci. Soc. Am. J.* 68, 1679-1688. <https://doi.org/10.2136/sssaj2004.1679>
- Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., Hole-Filled, S.R.T.M., 2008. for the globe Version 4. Available from the CGIAR/SGI SRTM 90m database: <http://srtm.csi.cgiar.org>.
- Jastrow, J.D., Miller, R.M., Lussenhop, J., 1998. Contributions of interacting biological mechanisms to soil aggregate stabilization in restored prairie. *Soil Biol. Biochem.* 30, 905-916.
- JAXA, 2019. ALOS Global Digital Surface Model "ALOS World 3D – 30m (AW3D30)". [https://doi.org/10.1016/S0038-0717\(97\)00207-1](https://doi.org/10.1016/S0038-0717(97)00207-1)
- Jiang, L., 2021. A fast and accurate circle detection algorithm based on random sampling. *Future Generation Computer Systems* 123, 245-256. <https://doi.org/10.1016/j.future.2021.05.010>
- Jones, E.J., Filippi, P., Wittig, R., Fajardo, M., Pino, V., McBratney, A.B., 2021. Mapping soil slaking index and assessing the impact of management in a mixed agricultural landscape. *Soil* 7(1), 33-46. <https://doi.org/10.5194/soil-7-33-2021>
- Kamamia, A.W., Vogel, C., Mwangi, H.M., Feger, K.H., Sang, J., Julich, S., 2021. Mapping soil aggregate stability using digital soil mapping: A case study of Ruiru reservoir catchment, Kenya. *Geoderma Regional* 24, p.e00355. <https://doi.org/10.1016/j.geodrs.2020.e00355>
- Kavdir, Y., Özcan, H., Ekinçi, H., Yüksel, O., Yiğini, Y., 2004. The influence of clay content, organic carbon and land use types on soil aggregate stability and tensile strength. *Turkish J. Agric. Forestry* 28(3), 155-162.
- Kemper, W.D., Rosenau, R.C., 1986. Aggregate stability and size distribution. *Methods of soil analysis: Part 1 Physical and mineralogical methods*, 5, 425-442. <https://doi.org/10.2136/sssabookser5.1.2ed.c17>
- Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling* 81, 401-418. <https://doi.org/10.1016/j.apm.2019.12.016>
- Khazaei, A., Mosaddeghi, M.R., Mahboubi, A.A., 2008. Impacts of test conditions, soil organic matter, clay and calcium carbonate contents on mean weight diameter and tensile strength of aggregates of some Hamedan soils. *JWSS-Isfahan Univ. Technology* 12(44), 123-134. <https://doi.org/10.1001.1.24763594.1387.12.44.10.5>

- Khosravani, P., Baghernejad, M., Moosavi, A.A., FallahShamsi, S.R., 2023. Digital mapping to extrapolate the selected soil fertility attributes in calcareous soils of a semiarid region in Iran. *J. Soils Sediments* 23(11), 4032-4054. <https://doi.org/10.1007/s11368-023-03548-1>
- Le Bissonnais, Y., 2016. Aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. *Eur. J. Soil Sci.* 67(1), 11-21. https://doi.org/10.1111/ejss.4_12311
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Liu, F., Rossiter, D.G., Song, X., Zhang, G.L., Wu, H., Zhao, Y., 2020. An approach for broad-scale predictive soil properties mapping in low-relief areas based on responses to solar radiation. *Soil Sci. Soc. Am. J.* 84(1), 144-162. <https://doi.org/10.1002/saj2.20025>
- Liu, Z., Wang, M., Zhou, J., Chen, Z., Xu, X., Zhu, Y., 2023. Soil aggregation is more important than mulching and nitrogen application in regulating soil organic carbon and total nitrogen in a semiarid calcareous soil. *Sci. Total Environ.* 854, 158790. <https://doi.org/10.1016/j.scitotenv.2022.158790>
- Ma, Z., Shi, Z., Zhou, Y., Xu, J., Yu, W., Yang, Y., 2017. A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai-Tibet Plateau with the effects of systematic anomalies removed. *Remote Sens. Environ.* 200, 378-395. <https://doi.org/10.1016/j.rse.2017.08.023>
- Mahjenabadi, V.A.J., Mousavi, S.R., Rahmani, A., Karami, A., Rahmani, H.A., Khavazi, K., *et al.*, 2022. Digital mapping of soil biological properties and wheat yield using remotely sensed, soil chemical data and machine learning approaches. *Computers Electronics Agric.* 197, 106978. <https://doi.org/10.1016/j.compag.2022.106978>
- Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154(1-2), 138-152. <https://doi.org/10.1016/j.geoderma.2009.10.007>
- Mansuy, N., Thiffault, E., Paré, D., Bernier, P., Guindon, L., Villemaire, P., *et al.*, 2014. Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. *Geoderma* 235, 59-73. <https://doi.org/10.1016/j.geoderma.2014.06.032>
- Martin, M.P., Dimassi, B., Román Dobarco, M., Guenet, B., Arrouays, D., Angers, D.A., *et al.*, 2021. Feasibility of the 4 per 1000 aspirational target for soil carbon: A case study for France. *Global Change Biol.* 27(11), 2458-2477. <https://doi.org/10.1111/gcb.15547>
- Mashalaba, L., Galleguillos, M., Seguel, O., Poblete-Olivares, J., 2020. Predicting spatial variability of selected soil properties using digital soil mapping in a rainfed vineyard of central Chile. *Geoderma Regional* 22, p.e00289. <https://doi.org/10.1016/j.geodrs.2020.e00289>
- Matinfar, H.R., Maghsodi, Z., Mousavi, S.R., Rahmani, A., 2021. Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale. *Catena* 202, 105258. <https://doi.org/10.1016/j.catena.2021.105258>
- McBratney, A.B., Santos, M.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117(1-2), 3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. *Advances Agronomy* 118, 1-47. <https://doi.org/10.1016/B978-0-12-405942-9.00001-3>
- Mishra, U., Hugelius, G., Shelef, E., Yang, Y., Strauss, J., Lupachev, A., *et al.*, 2021. Spatial heterogeneity and environmental predictors of permafrost region soil organic carbon stocks. *Science Advances* 7(9), p.eaaz5236. <https://doi.org/10.1126/sciadv.aaz5236>
- Mousavi, S.R., Sarmadian, F., Omid, M., Bogaert, P., 2022. Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran. *Measurement* 201, 111706. <https://doi.org/10.1016/j.measurement.2022.111706>
- Mousavi, S.R., Sarmadian, F., Angelini, M.E., Bogaert, P., Omid, M., 2023. Cause-effect relationships using structural equation modeling for soil properties in arid and semi-arid regions. *Catena* 232, 107392. <https://doi.org/10.1016/j.catena.2023.107392>
- Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2006. Use of the non-parametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Sci. Am. J.* 70(2), 327-336. <https://doi.org/10.2136/sssaj2005.0128>
- Nsabimana, G., Bao, Y., He, X., Nambajimana, J.D.D., Wang, M., Yang, L., *et al.*, 2020. Impacts of water level fluctuations on soil aggregate stability in the three gorges reservoir, China. *Sustainability* 12(21), 9107. <https://doi.org/10.3390/su12219107>
- Page, A.L., Miller, R.H., and Keeney, D.R., 1982. *Methodes of Soil Analysis Part2: Chemical and Biological Properties*, 2nd ed. Soil Sci. Am. J. Inc. Publisher Lected soil properties. *Soil Sci. Am. J.* 56, 557-561.
- Rahmani, A., Sarmadian, F., Arefi, H., 2022. Digital mapping of top-soil thickness and associated uncertainty using machine learning approach in some part of arid and semi-arid lands of Qazvin Plain. *Iranian J. Soil Water Res.* 53(3), 585-602. <https://doi.org/10.22059/IJSWR.2022.338007.669195>
- Resurreccion, A.C., Moldrup, P., Tuller, M., Ferré, T.P.A., Kawamoto, K., Komatsu, T., *et al.*, 2011. Relationship between specific surface area and the dry end of the water retention curve for soils with varying clay and organic carbon contents. *Water Res. Res.* 47(6). <https://doi.org/10.1029/2010WR010229>
- Rezaei, M., Mousavi, S.R., Rahmani, A., Zeraatpisheh, M., Rahmati, M., Pakparvar, M., *et al.*, 2023. Incorporating machine learning models and remote sensing to assess the spatial distribution of saturated hydraulic conductivity in a light-textured soil. *Computers Electronics Agric.* 209, 107821. <https://doi.org/10.1016/j.compag.2023.107821>
- Rostaminia, M., Rahmani, A., Mousavi, S.R., Taghizadeh-Mehrjardi, R., Maghsodi, Z., 2021. Spatial prediction of soil organic carbon stocks in an arid rangeland using machine learning algorithms. *Environ. Monitoring Assess.* 193, 1-17. <https://doi.org/10.1007/s10661-021-09543-8>
- Samaei F., Emami H., Lakzian A., 2022. Assessing soil quality of pasture and agriculture land uses in Shandiz country, north-eastern Iran. *Ecol. Indicators* 108974-108984. <https://doi.org/10.1016/j.ecolind.2022.108974>
- Shi, P., Castaldi, F., van Wesemael, B., Van Oost, K., 2020. Large-scale, high-resolution mapping of soil aggregate stability in croplands using APEX hyperspectral imagery. *Remote Sensing* 12(4), 666. <https://doi.org/10.3390/rs12040666>
- Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks* 19(2), 225-235. <https://doi.org/10.1016/j.neunet.2006.01.012>

- Soinne, H., Hyväluoma, J., Ketoja, E., Turtola, E., 2016. Relative importance of organic carbon, land use and moisture conditions for the aggregate stability of post-glacial clay soils. *Soil Till. Res.* 158, 1-9. <https://doi.org/10.1016/j.still.2015.10.014>
- Solomon, D., Fritzsche, F., Lehmann, J., Tekalign, M., Zech, W., 2002. Soil organic matter dynamics in the subhumid agroecosystems of the Ethiopian highlands: Evidence from natural ¹³C abundance and particle-size fractionation. *Soil Sci. Soc. Am. J.* 66(3), 969-978. <https://doi.org/10.2136/sssaj2002.9690>
- Song, X.D., Wu, H.Y., Ju, B., Liu, F., Yang, F., Li, D.C., *et al.*, 2020. Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma* 363, 114145. <https://doi.org/10.1016/j.geoderma.2019.114145>
- Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., *et al.*, 2019. Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrol. Earth System Sci.* 23(6), 2615-2635. <https://doi.org/10.5194/hess-23-2615-2019>
- Tajik, S., Ayoubi, S., Khajehali, J., Shataee, S., 2019a. Effects of tree species composition on soil properties and invertebrates in a deciduous forest. *Arabian J. Geosci.* 12(11), 368. <https://doi.org/10.1007/s12517-019-4532-8>
- Tajik, S., Ayoubi, S., Shirani, H., Zeraatpisheh, M., 2019b. Digital mapping of soil invertebrates using environmental attributes in a deciduous forest ecosystem. *Geoderma* 353, 252-263. <https://doi.org/10.1016/j.geoderma.2019.07.005>
- Tajik, S., Ayoubi, S., Zeraatpisheh, M., 2020. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. *Geoderma Regional* 20, p.e00256. <https://doi.org/10.1016/j.geodrs.2020.e00256>
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 266, 98-110. <https://doi.org/10.1016/j.geoderma.2015.12.003>
- Tang, X., Liu, S., Liu, J., Zhou, G., 2010. Effects of vegetation restoration and slope positions on soil aggregation and soil carbon accumulation on heavily eroded tropical land of Southern China. *J. Soils Sediments* 10, 505-513. <https://doi.org/10.1007/s11368-009-0122-9>
- Tejada M., Gonzalez J.L., 2008. Influence of two organic amendments on the soil physical properties, soil Losses, sediments and runoff water quality. *Geoderma* 145, 325- 334. <https://doi.org/10.1016/j.geoderma.2008.03.020>
- Wadoux, A.M.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews* 210, 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* 37(1), 29-38. <https://doi.org/10.1097/00010694-193401000-00003>
- Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., *et al.*, 2018. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Sci. Total Environ.* 630, 367-378. <https://doi.org/10.1016/j.scitotenv.2018.02.204>
- Wang, C., Zheng, M., Song, W., Wen, S., Wang, B., Zhu, C., *et al.*, 2017. Impact of 25 years of inorganic fertilization on diazotrophic abundance and community structure in an acidic soil in southern China. *Soil Biol. Biochem.* 113, 240-249. <https://doi.org/10.1016/j.soilbio.2017.06.019>
- Wang, C., Zhang, G., Zhu, P., Chen, S., Wan, Y., 2023. Spatial variation of soil functions affected by land use type and slope position in agricultural small watershed. *Catena* 225, 107029. <https://doi.org/10.1016/j.catena.2023.107029>
- Wang H., Guan D., Zhang R., Chen Y., Hu Y., Xiao H., 2014. Soil aggregates and organic carbon affected by the land use change from rice paddy to vegetable field. *J. Ecol. Eng.* 70, 211-260. <https://doi.org/10.1016/j.ecoleng.2014.05.027>
- Wang, Z., Shi, W., 2018. Robust variogram estimation combined with isometric log-ratio transformation for improved accuracy of soil particle-size fraction mapping. *Geoderma* 324, 56-66. <https://doi.org/10.1016/j.geoderma.2018.03.007>
- Weidhuner, A., Hanauer, A., Krausz, R., Crittenden, S.J., Gage, K., Sadeghpour, A., 2021. Tillage impacts on soil aggregation and aggregate-associated carbon and nitrogen after 49 years. *Soil Till. Res.* 208, 104878. <https://doi.org/10.1016/j.still.2020.104878>
- Wilding, L.P., Drees, L.R., 1978. Spatial Variability: A Pedologist's Viewpoint I. *Diversity Soils Tropics* 34, 1-12. <https://doi.org/10.2134/asaspecpub34.c1>
- World Health Organization, 2019. The state of food security and nutrition in the world 2019: safeguarding against economic slowdowns and downturns (Vol. 2019). Food Agric. Org. URL: http://www.fao.org/3/ca5162en/ca5_162en.pdf
- Wu, T., Luo, J., Dong, W., Sun, Y., Xia, L., Zhang, X., 2019. Geo-object-based soil organic matter mapping using machine learning algorithms with multi-source geo-spatial data. *IEEE J. Selected Topics Applied Earth Observations Remote Sensing* 12(4), 1091-1106. <https://doi.org/10.1109/JSTARS.2019.2902375>
- Yamaç, S.S., Şeker, C., Negiş, H., 2020. Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. *Agric. Water Manag.* 234, 106121. <https://doi.org/10.1016/j.agwat.2020.106121>
- Ye, L., Tan, W., Fang, L., Ji, L., Deng, H., 2018. Spatial analysis of soil aggregate stability in a small catchment of the Loess Plateau, China: I. Spatial variability. *Soil Till. Res.* 179, 71-81. <https://doi.org/10.1016/j.still.2018.01.012>
- Ye, L., Tan, W., Fang, L., Ji, L., 2019. Spatial analysis of soil aggregate stability in a small catchment of the Loess Plateau, China: II. Spatial prediction. *Soil Till. Res.* 192, 1-11. <https://doi.org/10.1016/j.still.2019.03.009>
- Yu, Y., Wan, M., Qian, J., Miao, D., Zhang, Z., Zhao, P., 2024. Feature selection for multi-label learning based on variable-degree multi-granulation decision-theoretic rough sets. *Int. J. Approximate Reasoning*, 169, 109181. <https://doi.org/10.1016/j.ijar.2024.109181>
- Zeraatpisheh, M., Khormali, F., 2012. Carbon stock and mineral factors controlling soil organic carbon in a climatic gradient, Golestan province. *J. Soil Sci. Plant Nutr.* 12(4), 637-654. <https://doi.org/10.4067/S0718-95162012005000022>
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* 338, 445-452. <https://doi.org/10.1016/j.geoderma.2018.09.006>
- Zeraatpisheh, M., Jafari, A., Bodaghabadi, M.B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., *et al.*, 2020.

- Conventional and digital soil mapping in Iran: Past, present, and future. *Catena* 188, 104424. <https://doi.org/10.1016/j.catena.2019.104424>
- Zeraatpisheh, M., Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M.R. Xu, M., 2021. Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. *Geoderma Regional* 27, p.e00440. <https://doi.org/10.1016/j.geodrs.2021.e00440>
- Zhao, D., Wang, J., Zhao, X., Triantafilis, J., 2022. Clay content mapping and uncertainty estimation using weighted model averaging. *Catena* 209, 105791. <https://doi.org/10.1016/j.catena.2021.105791>
- Zhao, Z.D., Zhao, M.S., Lu, H.L., Wang, S.H., Lu, Y.Y., 2023. Digital mapping of Soil pH based on machine learning combined with feature selection methods in East China. *Sustainability* 15(17), 12874. <https://doi.org/10.3390/su151712874>
- Zhou, Y., Chen, S., Zhu, A.X., Hu, B., Shi, Z., Li, Y., 2021. Revealing the scale-and location-specific controlling factors of soil organic carbon in Tibet. *Geoderma* 382, 114713. <https://doi.org/10.1016/j.geoderma.2020.114713>
- Zhu C., Li, X., Wang, C., Zhang, B., Li, B., 2024. Deep Learning-Based Coseismic Deformation Estimation from InSAR Interferograms. *IEEE Trans. Geoscience Remote Sensing*. In Press. Jan 22.
- Zolfaghari, A.A., Tirgar Soltani, M.T., Dyck, M., Weldeyohannes A., 2013. Comparison of K-nearest neighbor and artificial neural network methods for predicting cation exchange capacity of soil. 50th Anniversary Alberta Soil Science Workshop, Book of Abstracts, 48.