Agriculture, Environment & Society

Journal homepage: www.aes.uoz.ac.ir

Spatial modeling of soil saturation percentage using machine learning methods in Sistan plain

Younes Jamalzehi Samareh [a], Ali Shahriari [*b], Mohammad Reza Pahlavan-Rad [c], Alireza Ziaei Javid [d], Abolfazl Bameri [b]

[a] M.Sc Graduate of Soil Science, Department of Soil Science and Engineering, University of Zabol, Zabol, Iran
[b] Department of Soil Science and Engineering, University of Zabol, Zabol, Iran
[c] Soil and Water Research Department, Golestan Agricultural and Natural Resources Research and Education Center, AREEO, Gorgan, Iran
[d] Researcher, Devision of Soil Formation, Classification and Survey Researches, Soil and Water Research Institute, Karaj, Iran

ARTICLE INFO

Article history:
Received: 24 September 2024
Accepted: 21 February 2025
Available online: 1 June 2025

Keywords:
Deltaic soils
Hyper-arid region
Land management
Random forest

ABSTRACT

Soil maps are an urgent need for different land users and decision-makers. In recent years, attention to digital soil mapping has greatly increased, but most studies have focused on surface soil, even though land users are faced with the three-dimensional (3D) structure of soil. Saturation percentage (SP) is one of the physical attributes of soil moisture, which can be considered in land management, especially in the direction of soil water retention in connection with other attributes, in arid areas. Therefore, the present study was conducted with the aim of digital mapping of SP in 3D using some machine learning methods in the Sistan Plain, which is located on the Hirmand River delta in a hyper-arid region. To carry out this research, the information from 576 soil profiles located in the Sistan Plain was used and the percentage of saturated soil moisture was measured using the standard method at depths of 0-15, 15-30, 30-60, and 60-100 cm using the weighted average method. Random forest (RF), quantile regression forest (QRF), and cubist methods were used for spatial modeling. The results showed that the variables derived from remote sensing showed a significant correlation with the SP parameter only at the first and second depths, which were close to the ground surface, but the variables derived from DEM had a significant correlation at all depths. These variables were mainly related to alluvial activities, which had the greatest effect on soil changes in the studied area. Among the models, the RF method showed the best performance for spatial modeling of SP in all depths. The 3D modeling of the percentage of SP showed that the value of SP is the lowest in the south and medium in the middle of the area, and the highest in the north of the Sistan plain at the edge of the Hamoun wetlands. SP value is repeated with the same spatial trend, but the average value of SP increases from the surface to the depth. It seems that the changes in this attribute are in line with the 3D changes in the soil texture components in the region. Based on the results of the three-dimensional zoning of SP, it could be recommended that in the northern areas of the Sistan Plain, irrigation should be done with a longer time interval than in the southern regions for the same agricultural products. In the fields of natural resources, to manage vegetation and especially to deal with wind erosion, plants with shallow and deep roots in the northern regions, and trees and plants with deep roots in the southern regions can be considered. Machine learning methods, especially RF, can be effective in preparing digital and 3D maps of soil characteristics and can help different land users manage their land better.

Highlights

- RF model excels in 3D SP mapping in Sistan Plain, outperforming QRF and Cubist.
- SP lowest in south, highest in north near Hamoun wetlands, rises with depth.
- DEM variables key for SP at all depths; RS variables effective only at 0-30 cm.

* Corresponding author.
E-mail address: shahriari.ali@uoz.ac.ir
https://doi.org/10.22034/jelsa.2025.479520.1087

- SP aligns with soil texture changes, linked to alluvial activity in hyper-arid area.
- Suggests longer irrigation intervals in north, deep-root plants in south for management.

1. Introduction

As the foundation of terrestrial ecosystems, soil plays an important role in supporting biodiversity, agriculture, and ecosystem services (Montanarella and Panagos, 2021). It is essential to manage, utilize, and safeguard soil, while also comprehending its transformations at a landscape scale, to achieve sustainability (Mulder et al., 2011; Wulf et al., 2014). Soil studies are the main source for land use management and sustainable agriculture (Soil Science Division Staff, 2017). Soil is a controlling factor in many environmental processes such as greenhouse gas emissions, nitrate leaching, and plant and forest growth. Attributes such as acidity, salinity, texture, structure, and saturation percentage (SP) affect the physical and chemical behavior of the soil.

Saturated soil potential (SP) refers to the volume of water present in the soil when it is fully saturated. Specifically, SP is defined as the ratio of the weight of water needed to saturate the pore space to the weight of the dry soil. It serves as a critical indicator in soil hydroecological research. Additionally, SP is closely related to soil composition and can be employed as a parameter for estimating soil texture components, quantitatively assessing cation exchange capacity (CEC), and determining soil water holding capacit (Ahmad Aali et al., 2009). Also, soil SP reflects some physical properties of soil (Stivent and Khan, 1996). It is worth noting that soil SP indicates the amount of water availability for the plant, as well as the amount of movement of organic and inorganic solutes in the soil, and therefore, the processes of soil formation and the evolution of the profile affected by it (Ahmad Aali et al., 2009).

Identifying spatial changes of soil characteristics such as SP is essential for land management. Digital soil mapping (DSM) is a key method for assessing soil spatial changes and has been utilized for over two decades (Kidd et al., 2018; Kidd et al., 2020). Advances in computing technology have revolutionized soil science (Wadoux et al., 2020). DSM integrates field and laboratory methods with spatial and non-spatial inference systems to populate soil spatial information systems (IUSS, 2023). Recent years have seen a surge in DSM studies focusing on soil physical and chemical properties, driven by the increasing demand for quantitative soil data, advancements in statistical modeling and artificial intelligence, and improved access to environmental data for rapid soil mapping (Taghizadeh-Mehrjardi et al., 2021; Wadoux et al., 2020; Grunwald et al., 2012). Accurate soil datasets are essential for effective management of agriculture, water conservation, carbon stocks, and soil erosion (Adeniyi et al., 2024; Žížala et al., 2022). There are few studies in the direction of three-dimensional mapping of soil properties such as particle size distribution and soil texture class (Amirian Chekan et al., 2017; Dharumarajan and Hegde, 2022; Emami et al., 2024b), organic carbon (Jamshidi et al., 2019; Mousavi et al., 2022) and saltiness and salinity. (Filippi et al., 2020; Emami et al., 2024a) has been done. Researchers used the RF model for digital soil mapping at the province scale and showed that when the RF model was trained using conventional soil mapping (CSM), the accuracy of the resulting DSM was higher than the original CSM (Heung et al., 2022). Cao et al. (2023) used the RF model to predict the risk of cadmium contamination in the soil of an abandoned mine and showed that the RF model is an accurate and stable model for predicting the risk of toxic metal contamination. Emami et al. (2024b) Using the quantile regression forest (QRF) method, showed that environmental variables are the most important factors in predicting the distribution of soil texture components and early soil attributes such as SP. Also, the use of environmental variables as auxiliary variables is one of the important factors that influence DSM, and investigating the type of influence of these auxiliary variables can identify the type of soil, for a better understanding of soil development and preparing soil prediction maps is important and necessary (Duan et al., 2022). It is also possible to use different environmental variables that have constant application in the entire study area by considering the climatic and geographical conditions of the study area (Fan et al., 2022). Naimi Mardani et al. (2021), using synthetic soil image and machine learning made acceptable predictions using the Cubist model for sand, soil organic carbon and CCE, and the RF model for clay. They showed that the combination of high-quality RS data and DEM-derived variables can predict soil properties, and the use of RS data can reduce soil sampling costs and, as a result, soil mapping. The Sistan Plain is an alluvial plain located on the Hirmand River delta and has an extremely arid climate. SP is one of the easily available soil attributes that is related to the soil moisture level and water management in the soil. Therefore, paying attention to the 3D digital mapping of SP is very important for planning and management of water and soil in agricultural and environmental activities. So, the present study was conducted with the aim of digital mapping of the SP in 3D using some machine learning methods in the Sistan Plain

2. Materials and methods

2.1. Study area

Sistan Plain is located in the north of Sistan and Baluchistan province, southeast of Iran and southwest of the Asian continent with geographical coordinates of 61 degrees 10 minutes to 61 degrees 50 minutes of longitude and 30 degrees 18 minutes to 31 degrees 21 minutes of latitude. is (Figure 1). This deltaic plain is the result of the alluvial sediments of a river, which is known as a wide plain and originates from the Hirmand River in Afghanistan. The area of the studied area is about 218 thousand hectares (Siyasar et al., 2020; Piri, 2012). The Sistan Plain is a floodplain and does not have any orogenic activity and high elevations, and it has no special topography. The average height above sea level varies between 480-490 meters and the slope of the area varies

between 1-2% (Mirakzehi et al., 2018). According to the strategic characteristics of the Sistan plain, it has a hot, dry and desert climate, the average summer temperature is more than 40 degrees Celsius and the average winter temperature reaches 5 degrees above zero. In terms of humidity and temperature regime, the soil of the region is aridic and hyperthermic respectively. Its land use is divided into three types of agriculture, barren and unusable (salt marsh) and the predominant plant species are salt-tolerant plants and plants with deep roots (such as *Tamarix aphylla*), salt grass (*Salsola tomentosa*), heather (*Alhagi camelorum*), eucalyptus (*Eucalyptus camaldulensis*), Crete (D*esmostachya bipinnata*) (Mirakzehi et al., 2018; Delbari et al., 2019).

2.2. Sampling and laboratory methods

To estimate soil SP, the data from 576 soil profiles were used (Jamalzehi Samareh, 2022), and the method of determining the sampling points was based on the random supervised method (Figure 1). Soil SP tests have been measured in a standard way, using the weighted average method in Excel 2019 software to categorize the data and calculate the SP value at the depths of 0-15, 15-30, 30-60, and 60-100 cm was used. These depths were selected based on the method announced by the FAO organization to determine standard depths for studying and preparing soil maps (Bishop et al., 1999; Malone et al., 2009).

2.3. Environmental variables

Data obtained from remote sensing images can be used to obtain qualitative and quantitative information about soil properties and are an essential and very cost-effective data source for soil mapping (Agbu et al., 1990; Ben-Dor et al., 2009). In this study, Landsat 8 OLI images with a 27m resolution during whose cloud cover was less than 10% were downloaded from the EarthExplorer.gov site the sampling intervals, then the downloaded images were used for radiometric corrections with the Radiometric Calibration tool and atmospheric corrections with the FLAASH Atmospheric Correction tool in the software ENVI 5.3 was done. Subsequently, multiple indices were computed and derived using ArcGIS 10.4 software (Table 1). Additionally, a Digital Elevation Model (DEM) featuring a spatial resolution of 27 meters, obtained from the EarthExplorer.gov website, was utilized to extract environmental variables through SAGAGIS software (Table 2) and then in ArcGIS 10.4 software using the Extract Multi Values to Points tool, the variables of each point in the satellite images were extracted and using the Table to Excel variables were saved as a table in Excel 2019 software. After that, to determine the correlation between the studied attributes and environmental variables, Pearson's correlation coefficient was used in SPSS software, and the variables that showed a significant correlation with soil SP were included in the process of modeling and statistical analysis, and other variables were removed (Aksoy et al., 2012; Zeraatpishe et al., 2019; Shahriari et al., 2019). And then maps of each selected variable were prepared using ArcGIS 10.4 software (Wei et al., 2021; Bameri et al., 2015).

Initially, all environmental variables were incorporated into the modeling process. Subsequently, based on their relative importance, which ranges from 0 to 100%, variables with importance values below 10% were deemed insignificant and consequently excluded from the models. The remaining variables were utilized in the final models.

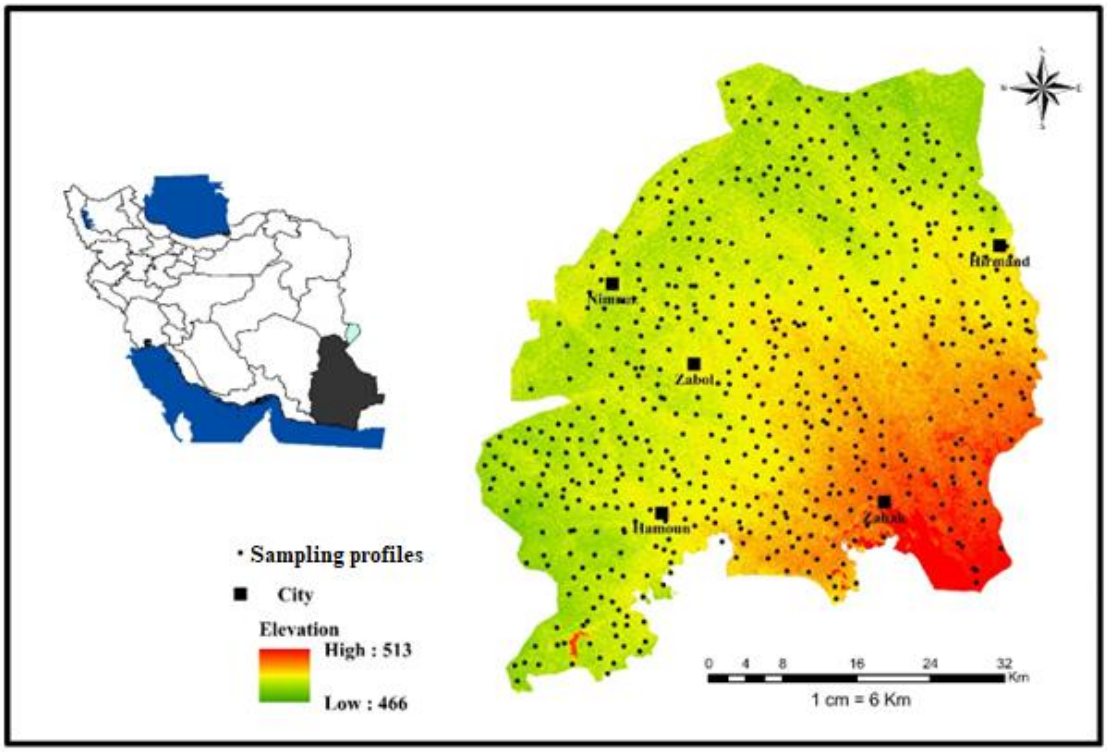

Figure 1. The location of the area and the studied soil profiles

2.4. Spatial modeling

The random forest (RF) method is a non-parametric model that builds multiple decision trees and combines them to create a more accurate classification. This method generates several trees and continuously classifies the data to find the best generalization point for each point and finally uses a voting mechanism in the forest to determine the output. This method is characterized by randomly selected samples and allows each tree in the forest to have similarities and differences (Cao et al., 2023; Shams et al., 2023).

The quantile regression forest (QRF) method as an advanced statistical modeling technique and a developed method of random forest algorithms is a flexible, non-linear, and non-parametric method that is very popular and useful as a powerful tool in machine learning methods (Gyamerah and Moyo, 2020; Athey et al., 2019; Wager and Athey, 2018). QRF has great potential to perform quantitative regressions on predictive distributions in various domains, as well as handle complex problems such as quantile regression and uncertainty estimation, and can provide users with a QRF algorithm that is widely implemented in R software (Gyamerah and Moyo, 2020; Meinshausen, 2017).

The Cubist method is a rule-based regression method that is quite effective in digital soil mapping (Zulfiqari and Abedi, 2019, Malone et al., 2016b). This algorithm is a tree model based on M5 theory (Quinlan, 1992) and prediction It divides the user into different subgroups based on "if-then" rules (Kuhn et al., 2016).

The main advantage of the Cubist method is to add multiple training committees and “reinforcement” so as to make the weights more balanced. a series of trees are produced to establish the Cubist model. The number of neighbors is used to modify the rule-based forecastsThe Cubist model employs a linear combination of the two models. The general conception about the Cubist regression model is described as follows: during the growth of a tree, many leaves and branches are grown.

The branches can be regarded as a series of “if-then” rules, while the terminal leaves can be regarded as an associated multivariate linear model. Assuming that a series of covariates comply with the condition of a rule, the associated model will be applied to calculate the predictive value. The Cubist model adds boosting with training committees (usually greater than one) which is similar to the method of “boosting” by sequentially developing a series of trees with adjusted weights. The number of neighbors in the Cubist model is applied to amend the rule-based prediction (Kuhn et al., 2016).

The key strength of the Cubist method lies in its use of multiple training committees and boosting techniques to achieve a more balanced distribution of weights. Its primary application is in analyzing large-scale databases that contain vast numbers of records and both numeric and nominal fields.

Each of the three models evaluates the importance of individual variables. During the modeling process, the values of each variable are randomly permuted while the values of the other variables remain unchanged. The difference between the Mean Squared Error (MSE) calculated from the permuted data and the original data provides a measure of variable importance. Variables that result in a relatively larger increase in MSE are considered more significant.

Initially, all environmental variables were included in the modeling process. However, based on their relative importance, which ranges from 0 to 100%, variables with importance values below 10% were deemed insignificant and subsequently removed from the models. The remaining variables were then used in the final models.

It is worth mentioning that to implement the desired models from the specialized packages "caret", "randomForest", "Cubist" and the "qrf" function was performed in R version 4.3.2 statistical software environment.

K-fold cross-validation was employed to assess model performance. Specifically, a 10-fold cross-validation approach was applied, in which the dataset was randomly divided into k=10 subsets (folds).

The model was trained using k=9 folds and validated on the remaining single fold, with accuracy metrics computed based on the test fold. This training and testing procedure was repeated k times, ensuring that each fold was used as the test set exactly once. The overall performance of the model was determined by averaging the accuracy metrics across all folds, and the results were reported accordingly.

The root mean square error (RMSE), mean error (ME), and correlation coefficient (R^2) were used to compare the models and choose the best model to obtain the dependent variables (saturated humidity). The R^2 value indicates the proportion of variation explained by the model, while the Root Mean Squared Error (RMSE) serves as a measure of prediction accuracy. The Mean Absolute Error (MAE), which is always a positive value, represents the average magnitude of errors and tends to exhibit a skewed distribution. Consequently, the closer the MAE value is to zero, the greater the accuracy of the method being evaluated.

Equation 1 was used to calculate RMSE:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathbb{Y}i - \mathbb{Y}j)^2} \qquad \text{Equation (1)}$$

where n is the total number of data, Yi is the measured value and Yj is the predicted value.

Equation 2 was used to calculate ME:

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}[\,\mathbb{Y}^*(\mathrm{xi}) - \mathbb{Y}(\mathrm{xi})] \qquad \text{Equation (2)}$$

where n is the total number of data, Y(xi) is the predicted value at the i-th point and Y^* (xi) is the calculated value at the i-th point.

Equation 3 was used to calculate R^2:

$$\mathbb{R}^2 = \frac{S_{xy}^2}{S_{xx}S_{xy}} \qquad \text{Equation (3)}$$

where x and y are the standard deviation of the variables and are in the form of Sx and Sy and its covariance is also denoted by COVxy and can be calculated from the relationship of R^2 (Metinfar et al., 2019).

Table 1. Environmental variables extracted from Landsat 8

Symbol of covariate	Description	Symbol of covariate	Description	Definition
SWI1	SWIR Band	DVI	Difference Vegetation Index	NIR – RED
SWIR2	SWIR Band	SR	Simple Ratio	NIR/Red
NIR	NIR Band	SLAVI	Specific Leaf Area Vegetation Index	NIR/(Red + SWIR 2)
coastal	Coastal Band	SAVI	Soil Adjusted Vegetation Index	(1 + L) * (NIR - RED) / (NIR + RED + L)
GEMI	-	MNDWI	Modified normalized difference water index	(Green -Swir)/ (Green + swir)
NBRI	Normalized burn ratio	EVI	Enhanced Vegetation Index	(NIR - RED) / (NIR + C1 * RED-C2 *BLUE + L)
CTVI	-	Blue	Blue band of Landsat-8	Wavelength of 0.450–0.515 μm
SATVI	soil-adjusted total vegetation index	Red	Red band of Landsat-8	Wavelength of 0.630–0.680 μm
NRVI	Normalized Ratio Vegetation Index	Green	Green band of Landsat-8	Wavelength of 0.525–0.600 μm
B1	Landsat OLI	EVI2	Enhanced Vegetation Index	(NIR - RED) / (NIR + C1 * RED-C2 *BLUE + L)
B2	Landsat OLI	NDVI	Normalized Difference Vegetation Index	(NIR-RED) / (NIR + RED)
B3	Landsat OLI	GNDVI	Green Normalized Difference Vegetation Index	(B5 - B3)/ (B5 + B3)
B4	Landsat OLI	NDWI	Normalized difference water index	(B3 - B5)/ (B3 + B5)
B5	Landsat OLI	NDWI2	Normalized difference water index	(B3 - B5)/ (B3 + B5)
B6	Landsat OLI	SAVI_1	Soil Adjusted Vegetation Index	(1 + L) * (NIR - RED) / (NIR + RED + L)
B7	Landsat OLI	MSAVI	Modified Soil-adjusted Vegetation Index	(1+L)(NIR-Red)/ (NIR+Red+L)
RVI	Ratio Vegetation Index	MSAVI2	Modified Soil-adjusted Vegetation Index	(2*NIR+1-sqrt ((2*NIR + 1)2-8*(NIR-Red)))/2

Table 2. Environmental variables extracted from DEM

Symbol of covariate	Description	References
Wetindex	Wetness index	-
Convergen	Convergence index	-
Diurnal_Anisotropic_ Heating	Diurnal anisotropic heating	Boettinger et al., (2008)
Overland Distflow	Overland flow distance to channel network	Boettinger et al., (2008)
Wind_Effect	The Wind Effect is a dimensionless index	Boettinger et al., (2008)
Drainage	Drainage Basins	Boettinger et al., (2008)
Cluster	Study area classification into groups	Boettinger et al., (2008)
Croscurve	Cross- sectional curvature	Boettinger et al., (2008)
Slope	Slope angle (%)	-
Catchment	Catchment area	-
Landforms	Landform	-
Distriverfinal	Distance from main river	-
Relative_Slope_ Position	Relative slope position	Boettinger et al., (2008)
Analytical _Hillshading	Analytical hillshading	Boettinger et al., (2008)
Channel network	Channel networks	Boettinger et al., (2008)
Flow_Path_ Length	Flow path length	Boettinger et al., (2008)
Effective_Air_ Flow_Heights	Effective air flow heights	Boettinger et al., (2008)
Horizontal_ Overland_ Flow_Distance	Horizontal overland flow distance	Boettinger et al., (2008)
Vertical Distance Channel	Vertical distance to channel network	Taghizadeh-Mehrjardi et al., (2021)
Valey depth	Depth of valley in meters	Rodiguez et al., (2002) Taghizadeh-Mehrjardi et al., (2021)
LSfacror	Multiple flow algorithms and help to accurately estimate current accumulation	Boehner and Selige, (2006) Taghizadeh-Mehrjardi et al., (2021)
Aspect	Compass direction of the maximum rate of change	Hom, (1981)

3. Results and discussion

3.1. Statistical analysis

Characteristics such as kurtosis and skewness of the data show that at all depths the SP data are normal to almost normal (Table 3). Also, SP values in all depths have the lowest coefficient of variation. Coefficient of variation (CV), which is a measure of relative variability, if 100% > CV ≥ 50% high variability, 50% > CV ≥ 21% moderate variability, and if CV ≤ 20% indicates low variability (Karimi Nezhad et al., 2015).

In this study, the coefficient of variation of SP was 50%-21%, which showed moderate variability. Also, the average value of SP showed that the first depth (0-15 cm) had the lowest average value (33.92 percent) and the fourth depth (60-100 cm) had the highest average (39.30 percent). Regarding SP changes in Sistan plain soil, limited studies have shown that in the Miankangi area, the average (46.67%) of this characteristic is higher than the average in the surface soil of the whole plain (Gholamalizadeh Ahangar et al., 2015; Hashemi et al., 2016).

According to the changes in soil texture components in the study area (Jamalzehi Samrah et al., 2021), it seems that the role of these attributes on SP changes at the surface and depth is significant.

3.2. Auxiliary variables

Figure 2 shows the distribution map of auxiliary variables with high importance in RF and QRF methods, including valley depth, channel network, distance from the river, evaluation, and normalized difference vegetation index (NDVI). Based on RF modeling, The valley depth variable was the most important, the channel network, the distance from the river, and the evaluation were less important, and the NDVI and vertical index were the least important (Figure 3-a). The channel network variable has the most importance, evaluation, valley depth, distance from the river, band 4, band 6, surface soil particle size index, band 7, band x43 and salinity index have medium importance, NDVI index has the least importance (Figure 3- b). The valley depth variable had the most importance, the channel network, distance from the river, and evaluation had medium importance, and the vertical and salinity index had the least importance (Figure 3-c). The valley depth variable had the most importance, evaluation, distance from the river and channel network had medium importance, and vertical and NDVI index had the least importance (Figure 3-d).

Table 3. Some statistical characteristics of SP

Depth (Cm)	Average	Median	Min	Max	Variance	Skewness	Kurtosis	Standard Deviation	CV
0-15	33.96	32.99	14	61.6	53.93	0.55	0.23	7.34	21.61
15-30	35.38	33.26	14	74.3	88.83	1.03	1.02	9.42	26.62
30-60	38.07	35.67	12	74.7	124.34	0.95	0.72	11.15	29.29
60-100	39.40	36.79	16	75.4	140.70	0.81	0.07	11.86	20.10

Based on QRF modeling, the variables of the channel network, evaluation, and distance from the river are the most important, valley depth variables, NDVI index, band 5, band 4, band 6 and band x43 are of medium importance, surface soil particle size index variables, band 7 and index Salinity was the least important (Figure 4-a). The variable of the channel network has the most importance, variables of evaluation, distance from the river, NDVI index, band 4, band 5, valley depth, band 6, band x43, and surface soil particle size index have medium importance, band 7 variables, and salinity index have the least importance were (Figure 4-b). The channel network variable had the most importance, the evaluation and distance from the river variables had medium importance, and the valley depth variable had the least importance (Figure 4-c). The channel network variable, evaluation has the most importance, the distance from the river has medium importance, and the valley depth variable has the least importance (Figure 4-d).

Emami et al. (2024b) used different auxiliary variables to model soil texture components using the QRF model and found that the valley depth variable acted as one of the best auxiliary variables for clay and sand modeling. Researchers considered the distance from the river and the network of canals (Pahlavan-Rad and Akbarimoghaddam, 2018) and Bands 4 and 7 (Shahriari et al., 2019) to be important auxiliary variables in the modeling of the surface soil texture of a part of the Sistan plain. Various researchers have found that elevation has played an important role in the preparation of soil maps as an auxiliary variable for modeling (Taghizadeh-Mehrjardi et al., 2021). Mirak zehi et al. (2018) showed in the spatial modeling of Sistan plain soil classes that the channel network, valley depth, elevation, and distance from the river are important auxiliary variables.

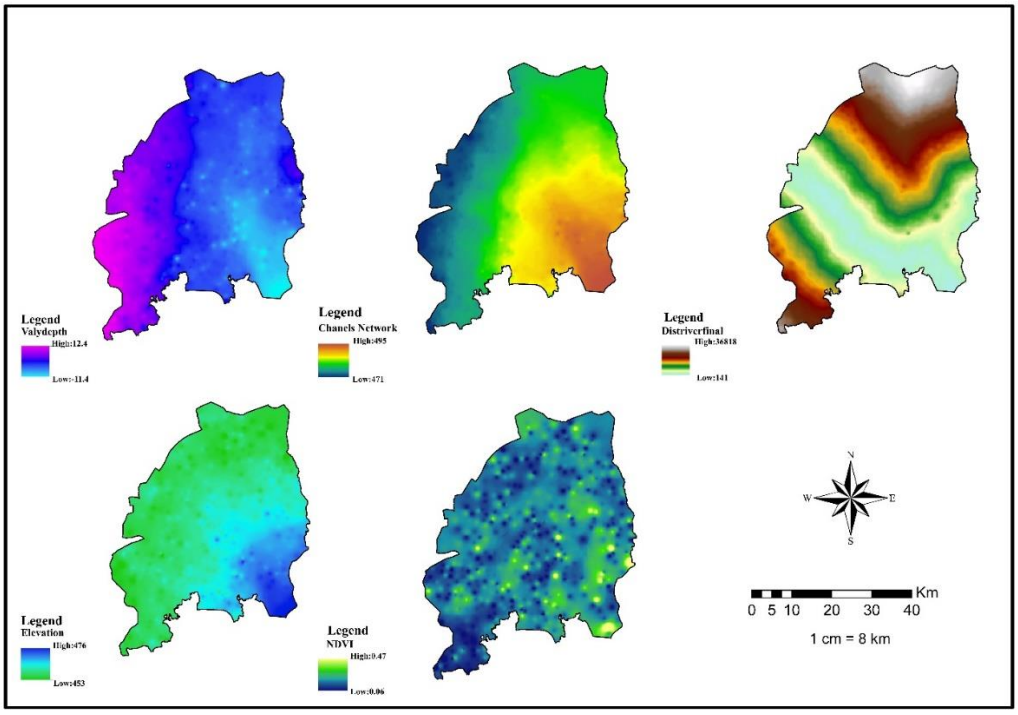

Figure 2. Distribution of auxiliary environmental variables in the study area

In general, it can be said that the variables derived from remote sensing were effective in modeling soil SP only in the first (0-15 cm) and second (15-30 cm) depths which were close to the land surface. Several studies also confirmed that in three-dimensional modeling of soil characteristics, the importance of auxiliary variables extracted from satellite images decreases with increasing depth (Amirian Chekan et al., 2017; Taghizadeh Mehrjardi et al., 2014).

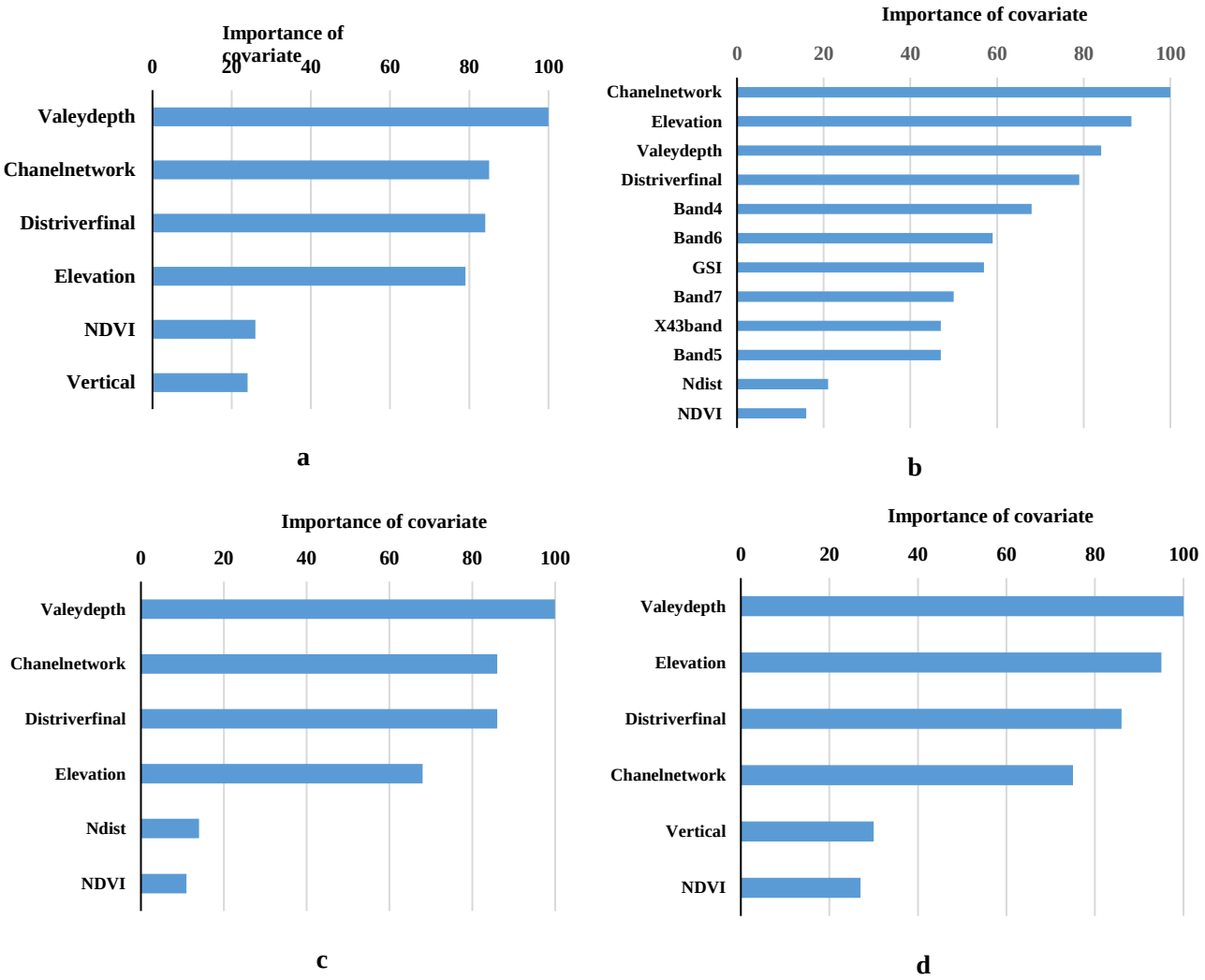

Figure 3. Importance of variables used in SP modeling with the RF model (a: 0-15 cm, b: 15-30 cm, c: 30-60 cm and d: 60-100 cm)

Also, several studies have stated that in areas with a low slope (less than 5%) due to the low variability of the variables from the DEM, other topographical parameters have less effect in predicting soil properties, especially on the surface (Amirian Chekan et al., 2017; Emami et al., 2024b). Contrary to these findings, in all studied depths, the variables derived from the DEM (terrestrial variables) were effective on SP in the area. In the third depth (30-60 cm) and the fourth depth (60-100 cm), only terrestrial variables were effective in modeling. Another point about these selected variables is that they are mainly related to alluvial activities (such as the depth of the valley, distance from the river, channel network, and elevation). The activity that has had the greatest effect on soil changes in the studied area (Mirakzehi et al., 2018; Pahlavan-rad and Akbarimoghaddam, 2018; Shahriari et al., 2019).

In other words, the selected environmental variables reflect the conditions and soil-forming factors in the Sistan Plain. On the other hand, it seems that for the physical attribute, SP, the variables derived from DEM show a higher correlation at all depths. The reason for the relationship between the variables extracted from DEM and physical characteristics such as soil texture components can be related to their effect on the vertical and lateral movement of soil particles through erosion and sedimentation (Emami et al., 2024b; Akpa et al., 2014).

3.3. Spatial modeling

Based on the highest value of R^2 and the lowest value of RMSE, the RF model for SP at depths of 0-15, 15-30, 30-60, and 60-100 was obtained as the best model for 3D spatial modeling in Sistan floodplain (Table 4). Of course, in the second depth, the QRF model has also performed better based on the MAE index. It should be mentioned that, in general, no significant difference was observed between the studied models. Also, Naimi Mardani et al.

(2021) showed that the RF model for the parameters of clay (R^2=0.48 and RMSE=6.02), silt (R^2=0.19 and RMSE=35.17), and sand (R^2=0 34.34 and RMSE=65.10) based on R^2 value and RMSE has a higher performance for predicting soil texture components. Hengl et al. (2015) predicted the spatial distribution of soil properties using the RF model and linear regression and reported that the RF model is more accurate in predicting soil properties than the linear regression model. Pahlavan-Rad and Akbarimoghaddam (2018) investigated the spatial distribution pattern of soil texture components and pH in a part of the agricultural lands of the Sistan floodplain in Zahak city by using the RF model and found that the RF model due to the large variation in the amount of soil texture in the plain Sistan flood reaches relatively high RMSE values for both soil texture and pH

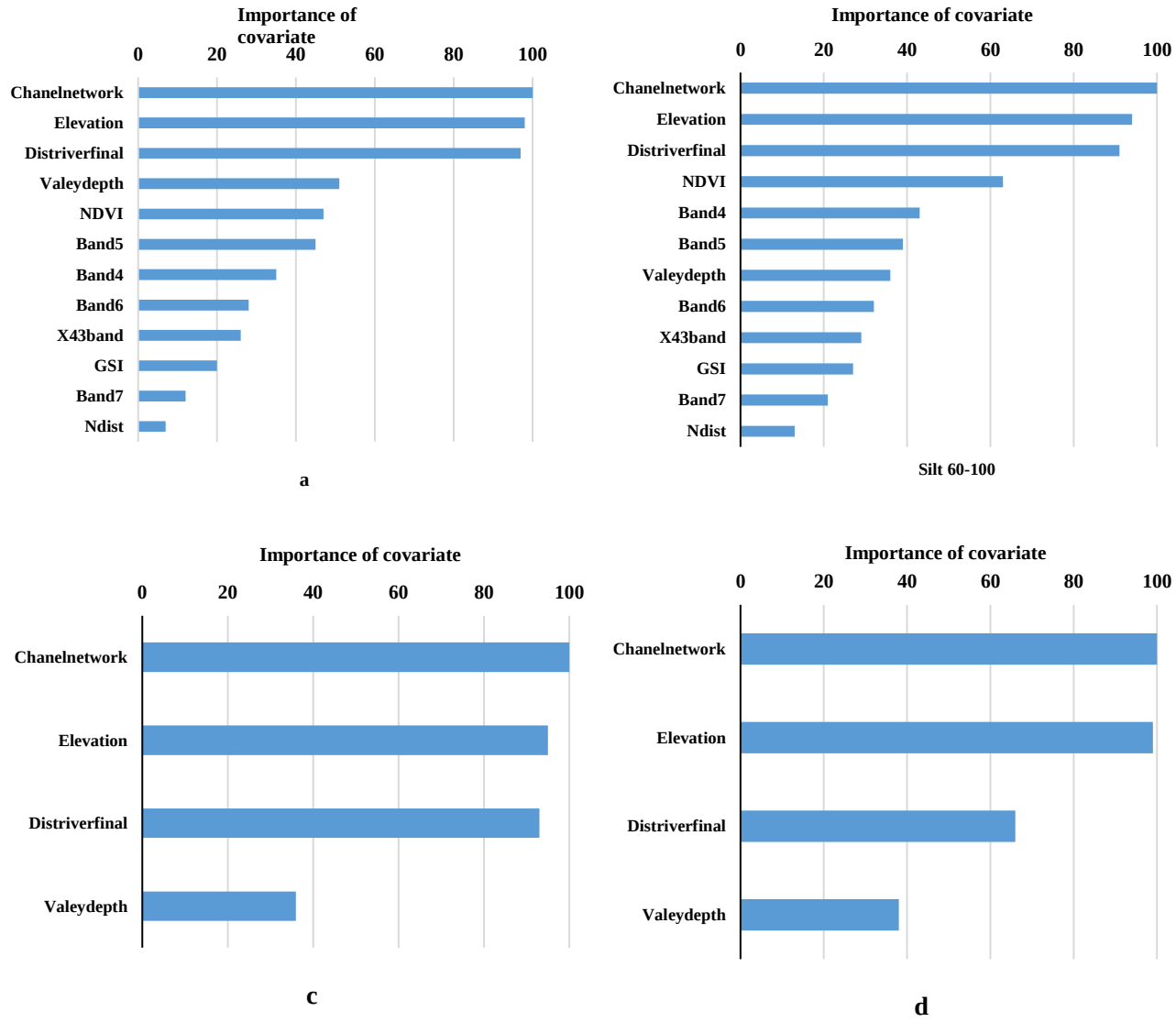

Figure 4. Importance of variables used in SP modeling with the QRF model (a: 0-15 cm, b: 15-30 cm, c: 30-60 cm and d: 60-100 cm)

Camera et al. (2017) using MLR models and RF, predicted soil properties including WRB soil groups, soil depth, and soil texture classes, and showed that the RF model compared to The MLR model performed better. Zhang et al. (2020) used the RF model to predict soil clay and showed that the RF model performs well in showing the changes from the surface to the depth and has a very good accuracy compared to other methods. Jena et al. (2023) used the RF model for digital mapping of soil texture components and their results showed that R^2 value for depths of 0-5, 5-15, 15-30, 30-60, 60-100, and -200 100 was equal to 0.30, 0.28, 0.21, 0.02, 0.02, 0.14, respectively, and in the first two depths, the accuracy of the model showed a good performance, and in the lower depths, the accuracy of the model decreased.

Lotfollahi et al. (2023) performed spatial modeling of soil texture components using the Global Soil Map and limited data and showed that the RF model with R^2 and RMSE values of 0.80 and 3.87 respectively for sand and 0.82 and 2.34 for clay and silt had the most accurate predictions of 0.85 and 2.89. Also, in this research, they showed that terrestrial-based environmental variables had a greater effect than remote-sensing variables.

Table 4. Soil SP spatial modeling results

Model	Depth (Cm)	Mtry	committees	Neighbors	RMSE	R^2	ME
RF	0-15	2	-	-	6.75	0.19	5.34
QRF		2	-	-	6.82	0.18	5.37
Cubist		-	1	9	6.99	0.17	5.47
RF	15-30	2	-	-	8.40	0.23	6.48
QRF		2	-	-	8.48	0.23	6.34
Cubist		-	10	0	8.45	0.19	6.46
RF	30-60	2	-	-	10.19	0.19	7.93
QRF		2	-	-	10.86	0.14	8.14
Cubist		-	1	0	10.54	0.15	8.03
RF	60-100	2	-	-	11.26	0.12	8.97
QRF		2	-	-	11.86	0.01	9.13
Cubist		-	1	9	11.52	0.11	9.09

The relatively high RMSE values observed in this study reflect the complexity of soil property variations and the diverse conditions influencing soil formation in floodplains (Wälder et al., 2008). Aeolian erosion and deposition are two significant processes contributing to increased spatial heterogeneity within the study area. Although challenging to incorporate, the inclusion of environmental variables related to aeolian processes could enhance the accuracy of the model.

3.4. 3D distribution of SP

As can be seen, the distribution of SP at depths of 0-15, 15-30, 30-60, and 60-100 is shown in Figure 5. The spatial changes of SP at the depth of 0-15 (first depth) has an increasing value in the northern parts and It has an average amount in the central parts of the Sistan plain and a smaller amount in the southern parts. In this depth, SP value in class more than 40% is 7.16% of lands, class 35-40% is 26.99% of lands, class 30-30% is 56.21% of lands, class 25-30 is 9.3% of lands and the class less than 25% included 0% of Sistan Plain lands. Hashemi et al. (2016) showed that the trend of SP changes in the surface soils of the Miankangi region of the Sistan Plain is similar to the pattern of distribution of silt and clay particles in the region. In other words, SP has a close and high relationship with soil texture components (Selmy et al., 2022).

According to research by Jamalzehi Samrah et al. (2021) on soil texture components in the Sistan region, a correlation analysis was done between these properties and SP at various depths (Table 5). The result showed a significant and positive correlation between silt and clay particles and SP in all depths. Conversely, a significant negative correlation was found between sand particles and SP in the studied area.

Table 5. Correlation analysis between SP and Soil texture components

	Depth	Clay%	Silt%	Sand%
SP%	0-15	0.581^{**}	0.491^{**}	-0.637^{**}
	15-30	0.631^{**}	0.442^{**}	-0.676^{**}
	30-60	0.647^{**}	0.379^{**}	-0.707^{**}
	60-100	0.655^{**}	0.324^{**}	-0.692^{**}

The significance level is marked with stars ($P < 0.01^{**}$)

The spatial changes of SP in the depth of 15-30 (second depth) have an increasing value in the northern parts, and it has an average value in the central parts of the plain and a small value in the southern parts. In this depth, SP value in class more than 40% is 22.60% of lands, class 35-40% is 20.89% of lands, class 30-30% is 50.50% of lands, class 25-30 is 73.5% of lands and the class included less than 25% of 0.01% of Sistan plain lands.

Shahriari et al. (2019) in the spatial modeling of the soil texture components of the Sistan plain at a depth of 0-30 cm showed that the amount of soil clay components is the highest in the northwestern and northern parts of the plain. Also, in the southern and central parts of the plain, which are adjacent to the Sistan River, there is more sand, and in the northwestern, northern, and western parts of the region, the amount of sand is the lowest. The amount of silt is the highest in the eastern, northeastern and western parts of the region, too. Also, Pahlavan-rad and Akbarimoghaddam (2018) in a study they conducted in the southeastern and southern parts of the Sistan plain investigated the changes in soil texture components at a depth of 0-30 cm and showed that due to the proximity to the Sistan river, the amount of sand in this area is high that the reason for that is the quick sedimentation of sand from suspension during floods. On the other hand, the high amount of sand in the area is caused by wind-blown sediments resulting from the 120-day winds that prevail in the area. These strong winds start from the beginning of June and end in the middle or end of September, when the water flow of the Hirmand River stops in the area from the north, and the northwest blows towards the south and causes the erosion of sand particles from the northern parts and the sedimentation and redistribution of sand particles in the southwestern and southern parts. Also, the amount of silt was high in the area around the Sistan River, and since the most important variable of these researchers in this study was "distance from the river", it is likely that the amount of clay is higher in these soils as a result of the slower settling of clay and silt particles from flood sediments above the river. in the region (Pahlavan-Rad and Akbarimoghaddam, 2018). Based on this, it seems that the amount of SP in this area is strongly influenced by the changes in the soil texture components and is in line with the changes in clay and silt parameters and has an opposite relationship with the sand parameter.

The spatial changes of SP in the depth of 30-60 (third depth) has an increasing value in the northern parts and has an average value in the central and southern parts of the plain. In this depth, SP value in class more than 40% is 27.33% of lands, class 40-35% is 34.49% of lands, class 30-30% is 31.23% of lands, class 25-30 is 1.01% of lands and the class less than 25% included 0% of Sistan Plain lands. The spatial distribution of SP in the depth of 60-100 (fourth depth) has an increasing value in the northern and central parts, and it has an average value in the southern parts of the plain. In this depth, SP value in class more than 40% is 39.31% of lands, class 35-40% is 37.03% of lands, class 30-30% is 23.93% of lands, class 25-30 is 0.72% of lands and the class less than 25% included 0% of Sistan Plain lands.

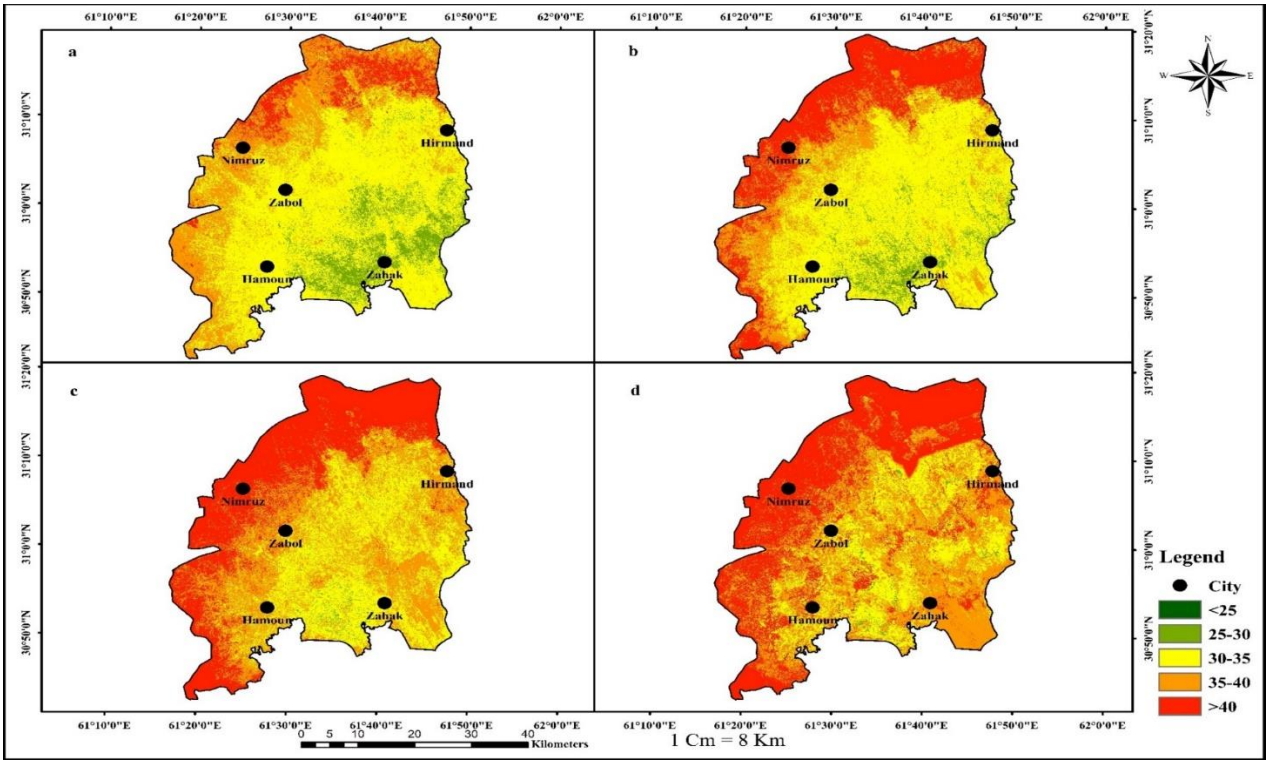

Figure 5. SP distribution at depths of a: 0-15 cm, b: 15-30 cm, c: 30-60 cm and d: 60-100 cm.

In general, the amount of SP increases from the surface to the depth of the soil in the Sistan plain. Jamalzehi Samrah et al. (2021) showed that with increasing soil depth, the proportion of fine soil components (silt and clay) increases with depth in the Sistan plain, and these changes are constant from 40 cm to 100 cm deep. Therefore, as mentioned, the changes in soil SP in the Sistan plain are in line with the changes in the soil texture components of the region (Hashemi et al., 2015; Selmy et al., 2022). On the other hand, the soil SP 3D changes are in line with the changes in the soil texture of the Sistan plain, and these changes are strongly influenced by the interactions of the prevailing alluvial and aeolian activities in the studied area, that these activities cause complexity in the relative changes of the soil characteristics of the region (Mirakzehi et al., 2018; Pahlavan-rad and Akbarimoghaddam, 2018; Shahriari et al., 2019).

Based on the results obtained regarding the three-dimensional zoning of the percentage of saturated soil moisture, it seems that in the northern areas of the Sistan plain, due to the high level of this parameter, the irrigation cycle can be done with a longer time interval than in the southern areas of the plain for agricultural products. Therefore, agricultural products or plants with more drought resistance can be recommended for the northern areas of the plain. Also, in the northern regions, both plants with shallow and deep roots can be considered to manage vegetation and deal with wind erosion, in the fields of natural resources, and in the southern regions of the plain, more trees and plants with deep roots can be used for this purpose.

4. Conclusion

3D spatial modeling of soil SP in dry areas is one of the important aspects of soil mapping that plays an important role in land management strategies. For the spatial modeling of this attribute, different methods are used, including random forest, quintile regression forest, and cubist. The results showed that the value of SP in the south was the lowest values and in the middle of the Sistan plain was the average values and in the north of the plain at the edge of the Hamoun wetlands had the highest values and the SP changed from the surface to the depth with the same spatial trend in the studied layers, and its average value increases from the surface to the depth. It seems that the changes are in line with the three-dimensional changes of the soil texture components in the region. The random forest method had the best performance in SP estimation and the environmental variables derived from satellite images and DEM (terrestrial variables) at two upper depths

(0-15 and 15-30 cm) and the terrestrial variables at all depths showed a good correlation with SP. The relationship between selected environmental variables and alluvial processes was significant. In general, the findings emphasize the importance of integrating environmental variables that reflect soil conditions and their influencing factors in conjunction with advanced modeling techniques such as RF to prepare and produce soil characteristic maps with high accuracy. For future studies, it is recommended to use new machine learning models in this regard.

References

Adeniyi, O.D., Bature, H., & Mearker, M. A. (2024). Systematic Review on Digital Soil Mapping Approaches in Lowland Areas. Land, 13, 379. https://doi.org/10.3390/land13030379.

Agbu, P. A., Fehrenbacher, D. J., & Jansen, I. J. (1990). Soil property relationships with SPOT satellite digital data in east central Illinois. *Soil Science Society of America Journal*, *54*(3), 807-812. https://doi.org/10.2136/sssaj1990.03615995005400030031x

Ahmad Aali, K., Parsinejad, M., & Rahmani, B. (2009). Estimation of Saturation Percentage of Soil Using Multiple Regression,ANN, and ANFIS Techniques. *Computer and Information Science*, *2*(3), 127-136

Akpa, S. I. C., Odeh, I. O. A. & Bishop, T. F. A. (2014). Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal, 78*. 1953-1966. https://doi.org/10.2136/sssaj2014.05.0202

Aksoy, E., Panagos, P., & Montanarella, L. (2012). Spatial prediction of soil organic carbon of Crete by using geostatistics. In: Minasny, B., Malone, B.P., McBratney, A.B. (Eds.), Digital Soil Assessments and Beyond. *CRC Press, London*, 149–159.

Amirian Chekan, A., Taghizadeh Mehrjerdi, R., Sarmadian, F., & Heidary, A. (2017). Three-dimensional mapping of soil texture using spline depth functions and artificial neural networks. *Iranian Journal of Soil and Water Research*, *48*(1). 113-123. [In Persian]

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. https://doi.org/10.1214/18-aos1709.

Bameri, A., Khormali, F., Kiani, F., & Dehghani, A. A. (2015). Spatial variability of soil organic carbon in different hillslope positions in Toshan area, Golestan Province, Iran: Geostatistical approaches. *Journal of Mountain Science*, *12*(6), 1422-1433. https://doi.org/10.1007/s11629-014-3213-z

Ben-Dor, E., Chabrillat, S., Dematt, J., Taylor, G., Hill, J., Whiting, M., & Sommer, S. (2009). Using imaging spectroscopy to study soil properties. *Remote Sens. Environ. 113* (1), 38-55. Imaging Spectroscopy Special Issue. https://doi.org/10.1016/j.rse.2008.09.019OUCI+2

Bishop, T. F. A., McBratney, A. B. & Laslett, G. M. (1999). Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma, 91*, 27-45. https://doi.org/10.1016/s0016-7061(99)00003-8

Boehner, J., & Selige, T. (2006). Spatial prediction of soil attributes using terrain analysis and climate regionalisation. *Gottinger Geographische Abhandlungen*, *115*, 13-28

Boettinger, J. L., Ramsey, R. D., Bodily, J. M., Cole, N. J., Kienast-Brown, S., Nield, S. J., ... & Stum, A. K. (2008). Landsat spectral data for digital soil mapping. In *Digital soil mapping with limited data* (pp. 193-202). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-8592-5_16

Camera, C., C., Zomeni, Z., Noller, J. S., Zissimos, A. M., Christoforou, I. C., & Bruggeman, A. (2017). A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma*, *285*, 35-49. https://doi.org/10.1016/j.geoderma.2016.09.019

Cao, J., Guo, Z., Lv, Y., Xu, M., Huang, C., & Liang, H. (2023). Pollution risk prediction for cadmium in soil from an abandoned mine based on random forest model. *International Journal of Environmental Research and Public Health*, *20*(6), 5097. https://doi.org/10.3390/ijerph20065097

Delbari, M., Afrasiab, P., Gharabaghi, B., Amiri, M., & Salehian, A. (2019). Spatial variability analysis and mapping of soil physical and chemical attributes in a salt-affected soil. *Arabian Journal of Geosciences*, *12*(3), 68-86. https://doi.org/10.1007/s12517-018-4207-x.

Dharumarajan, S., & Hegde, R. (2022). Digital mapping of soil texture classes using Random Forest classification algorithm. *Soil Use and Management*, *38*(1), 135-149. https://doi.org/10.1111/sum.12668

Duan, M., Guo, Z., Zhang, X., & Wang, C. (2022). Influences of different environmental covariates on county-scale soil type identification using remote sensing images. *Ecological Indicators*, *139*, 108951. https://doi.org/10.1016/j.ecolind.2022.108951

Emami, M., Khormali, F., Pahlavan-Rad, M. R., & Ebrahimi, S. (2024). Digital modeling of surface and subsurface soil salinity in Golestan province, Iran. *Geoderma Regional*, *37*, e00800. https://doi.org/10.1016/j.geodrs.2024.e00800.

Emami, M., Khormali, F., Pahlavan-Rad, M. R., & Ebrahimi, S. (2024b). Preparation of three-dimensional maps of soil particle size fraction by combining quantile regression forest algorithm and spline functions in the north of Golestan province. Iran., *Iranian Journal of Soil and Water Research*, *55* (1), 51-68. [In Persian] https://doi.org/10.22059/ijswr.2023.366978.669594

Fan, N. Q., Zhao, F. H., Zhu, L. J., Qin, C. Z., & Zhu, A. X. (2022). Digital soil mapping with adaptive consideration of the applicability of environmental covariates over large areas. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 102986. https://doi.org/10.1016/j.jag.2022.102986

Filippi, P., Jones, E., & Bishop, T.F. (2020). Catchment-scale 3D mapping of depth to soil sodicity constraints through combining public and on-farm soil databases – A potential tool for on-farm management. *Geoderma*,

374, 114396. https://doi.org/10.1016/j.geoderma.2020.114396

Gholamalizadeh Ahangar, A., Sarani, F., Hashemi, M., & Shabani, A. (2015). Comparison of linear regression methods, geostatistical and artificial neural network modeling of organic carbon in dry land of Sistan plain. *Water and Soil*, *28*(6), 1250-1260. [In Persian] https://doi.org/10.22067/jsw.v0i0.32714

Grunwald, S., Thompson, J. A., Minasny, B., & Boettinger, J. L. (2012). Digital soil mapping in a changing world. In *Digital Soil Assessments and Beyond* (pp. 301-306). CRC Press London. https://doi.org/10.1201/b12728-60

Gyamerah, S. A., & Moyo, E. (2020). Long-Term Exchange Rate Probability Density Forecasting Using Gaussian Kernel and Quantile Random Forest. *Complexity*, *2020*(1), 1972962. https://doi.org/10.1155/2020/1972962

Hashemi, M. , Gholamalizadeh Ahangar, A. , Bameri, A. , Sarani, F. & Hejazizadeh, A. (2016). Survey and Zoning of Soil Physical and Chemical Properties Using Geostatistical Methods in GIS (Case Study: Miankangi Region in Sistan). *Water and Soil*, *30*(2), 443-458. [In Persian]. https://doi.org/10.22067/jsw.v30i2.25950

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., ... & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS one*, *10*(6), e0125814. https://doi.org/10.1371/journal.pone.0125814

Heung, B., Bulmer, C. E., Schmidt, M. G., & Zhang, J. (2022). Provincial-scale digital soil mapping using a random forest approach for British Columbia. *Canadian Journal of Soil Science*, *102*(03), 597-620. http://dx.doi.org/10.1139/cjss-2021-0090

Hom, B. K. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, *69*(1), 14-47. https://doi.org/10.1109/proc.1981.11918

IUSS. 7th global digital soil mapping workshop (2016). Available online: https://projects.au.dk/digitalsoilmapping// (accessed on 1st August 2023).

Jamalzehi Samrah, Y. (2022). *Three-dimensional spatial modeling texture components and saturation percentage of soil in the Sistan plain.* MSc's thesis. Faculty of Water and Soil, University of Zabol. 186 p.

Jamalzehi Samareh, Y., Shahriari, A., Pahlavan-Rad, M., Ziaei Javaid, A. & Bameri, A. (2021). *Preparation of three-dimensional maps of the size of soil particles in the floodplain of Sistan* [Conference presentation]. The Seventeenth Iran Soil Science Congress and 4th National Farm Water Management Conference, Soil and Water Research Institute, Karaj, Iran. [In Persian]

Jamshidi, M., Delavar, M. A., Taghizadehe-Mehrjerdi, R., & Brungard, C. (2019). Evaluating digital soil mapping approaches for 3D mapping of soil organic carbon. *Iranian Journal of Soil Research*, *33*(2), 227–239. [In Persian]. https://doi.org/10.22092/ijsr.2019.119764

Jena, R. K., Moharana, P. C., Dharumarajan, S., Sharma, G. K., Ray, P., Deb Roy, P., Ghosh, D., Das, B., Alsuhaibani, A. M., & Gaber, A. (2023). Spatial prediction of soil particle-size fractions using digital soil mapping in the North Eastern region of India. *Land*, *12*(7), 1295. https://doi.org/10.3390/land12071295

Karimi Nezhad, M. T., Tabatabaii, S. M., & Gholami, A. (2015). Geochemical assessment of steel smelter-impacted urban soils, Ahvaz, Iran. *Journal of Geochemical Exploration*, *152*, 91–109. https://doi.org/10.1016/j.gexplo.2015.01.005

Kidd, D., Searle, R., Grundy, M., McBratney, A., Robinson, N., O'Brien, L., Zund, P., Arrouays, D., Thomas, M., Padarian, J., Jones, E., Bennett, J. M., Minasny, B., Holmes, K., Malone, B. P., Liddicoat, C., Meier, E. A., Stockmann, U., Wilson, P., Wilford, J., Payne, J., Ringrose-Voase, A., Slater, B., Odgers, N., Gray, J., van Gool, D., Andrews, K., Harms, B., Stower, L., & Triantafilis, J. (2020). Operationalising digital soil mapping – Lessons from Australia. *Geoderma Regional*, *23*, e00335. https://doi.org/10.1016/j.geodrs.2020.e00335

Kidd, D., Searle, R., & Wilson, P. (2018). Digital soil mapping: application, opportunity and challenges. *Soil Science Policy Journal* (1), 35–40. Available at https://www.soilscienceaustralia.org.au/publications/soil-policy-journal/

Kuhn, M. K., Weston, S., Keefer, C., & Coulter, N. (2016). *Cubist models for regression* [Computer software manual]. https://cran.r-project.org/package=Cubist

Lotfollahi, L., Delavar, M. A., Biswas, A., Jamshidi, M., Taghizadeh-Mehrjardi, R., & Fatehi, S. H. (2023). Modeling the spatial distribution of sand, silt, and clay particles based on GlobalSoilMap and limited data. *DESERT*, *28*(2), 243–263.

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*, 138-152. https://doi.org/10.1016/j.geoderma.2009.10.007

Malone, B. P., Minasny, B., & McBratney, A. B. (2016). *Using R for digital soil mapping* (pp. 133–136). Springer. https://doi.org/10.1007/978-3-319-44327-0

Meinshausen, N. (2017). *quantregForest: Quantile regression forests* [Computer software manual]. CRAN.

Metinfar, H., Maqsoodi, Z., Mousavi, S., & Jalali, M. (2019). Evaluation of machine learning methods in digital organic carbon mapping of agricultural soils (part of Khorram Abad plain. *Water and Soil Magazine*, *24*(4), 327–343. [In Persian]

Mirakzehi, K., Pahlavan-Rad, M. R., Shahriari, A., & Bameri, A. (2018). Digital soil mapping of deltaic soils: A case of study from Hirmand (Helmand) river delta. *Geoderma*, *313*, 233–240. https://doi.org/10.1016/j.geoderma.2017.10.039

Montanarella, L., & Panagos, P. (2021). The relevance of sustainable soil management within the European Green Deal. *Land Use Policy*, *100*, 104950. https://doi.org/10.1016/j.landusepol.2020.104950

Mousavi, S. R., Sarmadian, F., Omid, M. K., & 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

Mulder, V., de Bruin, S., Schaepman, M., & Mayr, T. (2011). The use of remote sensing in soil and terrain mapping - a review. *Geoderma, 162*(1–2), 1–19. https://doi.org/10.1016/j.geoderma.2011.03.021

Naimi mardani, S., Ayoubi, S., Demattê, J., Zeraatpisheh, M., Amorim, M. T., & Oliveira Mello, F. A. (2021). Spatial prediction of soil surface properties in an arid region using synthetic soil image and machine learning. *Geocarto International*, 1–25. https://doi.org/10.1080/10106049.2021.1996639

Pahlavan-Rad, M. R., & Akbarimoghaddam, A. (2018). Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *Catena, 160*, 275–281. https://doi.org/10.1016/j.catena.2017.09.004

Piri, H. (2012). Assessment of computational methods of estimation of potential evapotranspiration using lysimeter data (Case study: Sistan Plain). *Journal of Irrigation and Water Engineering, 3*(9), 50–62. [In Persian]

Quinlan, J. R. (1992). Learning with continuous classes. In *Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence* (pp. 343–348). World Scientific.

Rodiguez, F., Maire, E., Courjault-Rade, D., & Darrozes, J. (2002). The Black Top Hat function applied to a DEM: A tool to estimate recent incision in a mountainous watershed. *Geophysical Research Letters, 29*(1), 91–94. https://doi.org/10.1029/2001gl013484

Selmy, S. A., Abd El-Aziz, S., El-Desoky, A., & El-Sayed, M. A. (2022). Characterizing, predicting, and mapping of soil spatial variability in Gharb El-Mawhoub area of Dakhla Oasis using geostatistics and GIS approaches. *Journal of the Saudi Society of Agricultural Sciences, 21*(4), 383–396. https://doi.org/10.1016/j.jssas.2021.05.004

Shahriari, M., Delbari, M., Afrasiab, P., & Pahlavan-Rad, M. R. (2019). Predicting regional spatial distribution of soil texture in floodplains using remote sensing data: A case of southeastern Iran. *Catena, 182*, 104–149. https://doi.org/10.1016/j.catena.2019.104149

Shams, M. Y., Tarek, Z., Elshewey, A. M., Hany, M., Darwish, A., & Hassanien, A. E. (2023). A machine learning based model for predicting temperature under the effects of climate change. In A. E. Hassanien & A. Darwish (Eds.), *The power of data: Driving climate change with data science and artificial intelligence innovations* (pp. 61–81). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-22457-4_4

Siyasar, H., Honar, T., & Abdolahipour, M. (2020). Comparing of generalized linear models, random forest and gradient boosting trees in estimation of reference crop evapotranspiration (Case study: The Sistan Plain). *JWSS, 23*(4), 395–410. [In Persian] https://doi.org/10.47176/jwss.23.4.40631

Soil Science Division Staff. (2017). *Soil survey manual* (C. Ditzler, K. Scheffe, & H. C. Monger, Eds.) (USDA Handbook 18). Government Printing Office.

Stivent, G. A., & Khan, M. A. (1996). Saturation percentage as a measure of soil texture in the lower Indus Basin. *Journal of Soil Science, 17*(1), 255–263. https://doi.org/10.1111/j.1365-2389.1966.tb01471.x

Taghizadeh Mehrjardi, R., Minasny, B., Sarmadian, F., & Malone, B. P. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma, 213*, 15–28. https://doi.org/10.1016/j.geoderma.2013.07.027

Taghizadeh-Mehrjardi, R., Emadi, M., Cherati, A., Heung, B., Mosavi, A., & Scholten, T. (2021). Bio-inspired hybridization of artificial neural networks: An application for mapping the spatial distribution of soil texture fractions. *Remote Sensing*, Article 1–23. https://doi.org/10.3390/rs13040612

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

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Wei, Y., Ding, J., Yang, S., Yang, X., & Wang, F. (2021). Comparisons of random forest and stochastic gradient treeboost algorithms for mapping soil electrical conductivity with multiple subsets using Landsat OLI and DEM/GIS-based data at a type oasis in Xinjiang, China. *European Journal of Remote Sensing, 54*(1), 158–181. https://doi.org/10.1080/22797254.2021.1888657

Wulf, H., Mulder, T., Schaepman, M., Keller, A., & Jörg, P. (2014). *Remote sensing of soils* (Tech. Rep. 00.0338.PZ/435-0501). University of Zurich. https://doi.org/10.5167/uzh-109992

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.2018.11.011

Zhang, Y., Ji, W., Saurette, D. D., Easher, T. H., Li, H., Shi, Z., Adamchuk, V. I., & Biswas, A. (2020). Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging. *Geoderma, 366*, 114253. https://doi.org/10.1016/j.geoderma.2020.114253

Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Rojas, J. R., Penížek, V., & Zádorová, T. (2022). High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. *Catena, 212*, 106024. https://doi.org/10.1016/j.catena.2022.106024

Zulfiqari, F., & Abedi, Gh. (2019). Modeling the effective factors in the concentration of dust particles in the air using the method CUBIST. In *International Conference on Dust in Southwest Asia*, Zabol. [In Persian]