DESERT 2021, 26(2): 173-186
DOI: 10.22059/jdesert.2021.298777.1006768

RESEARCH PAPER

Identifying the determinant characteristics influencing soil compactibility indices using neural networks and path analysis Running title: Investigation of Soil Compaction by ANN and PA

Hossain Shirani[1,*], Mohamad Reza Mosaddeghi[2], Naghme Rafinejad[1], Somayeh Sadr[3], Hossain Dashti[4], Hormozd Naghavi[5]

[1] Department of Soil Science, College of Agriculture, Vali-e-Asr University of Rafsanjan, Kerman, Iran
[2] Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, Iran
[3] College of Agriculture, Payame Noor University of Kerman, Iran
[4] Department of Plant Breeding, College of Agriculture, Vali-e-Asr University of Rafsanjan, Kerman, Iran
[5] Department of Water and Soil Science, Agricultural and Natural Resources Research Center, Agricultural Research, Training and Promotion Organization, Ministry of Agriculture, of Kerman Province, Iran

Received: 02 March 2020, Revised: 21 February 2021, Accepted: 08 August 2021

Abstract
Soil compactibility could be quantified via different indices, such as maximum dry bulk density (BD_{max}) and critical water content ($\theta_{critical}$) in a compaction test. The objective of this study was to determine the soil properties influencing soil compactibility by evaluating pedotranfer functions (PTFs) with respect to their accuracy and function for the prediction of BD_{max} and $\theta_{critical}$ using linear regression and artificial neural network (ANN) methods. To this end, 100 soil samples were collected from the topsoil (0−30 cm) of arable and virgin lands in southeast of Iran. Primary particle size distribution, gypsum, Calcium Carbonate Equivalent (CCE), organic matter (OM) contents, and natural bulk density were used as predictors. Two PTFs were developed using linear multiple stepwise regression: a PTF that estimates BD_{max} with clay and sand contents and natural bulk density as predictors ($R^2 = 0.45$), and another PTF for the estimation of $\theta_{critical}$ employing clay and gypsum contents as predictors ($R^2 = 0.51$). Furthermore, an attempt was made to construct PTFs for the prediction of the BD_{max} and $\theta_{critical}$ utilizing ANNs. High prediction efficiencies were achieved through the ANN models. Generally, when all of the easily available soil properties were included as predictors, the ANN models for the θcritical and BDmax as compared with the results of linear regression method obtained much more accurate estimations. Sensitivity analysis performed in this study via Hill method (shirani, 2017) showed that the most important variable in BD_{max} prediction-using ANNs is the natural bulk density ($BD_{natural}$), followed by sand and clay, CCE, and gypsum contents. The highest sensitivity to clay content belonged to the $\theta_{critical}$ and the lowest sensitivity to OM content was observed in the studied soils.

Keywords: Pedotransfer functions, Linear regression, Maximum dry bulk density, Critical water content, Proctor compaction test

Introduction

Soil compaction is a process of soil densification due to external mechanical forces which increases bulk density, reduces porosity and air/water permeability, and consequently lowers soil

* Corresponding author e-mail: shirani@vru.ac.ir

physical quality (Soane and Van Ouwerkerk, 2013). In agricultural science, low and moderate soil compaction is usually beneficial in terms of good soil-seed contacts and mechanical confinement of plant roots in the soil. Therefore, studying soil compaction in agricultural engineering is complicated given the facts that agricultural soils are structured, unsaturated, and heterogeneous, and soil mechanics principles should be adapted to agricultural situations with causation (Mosaddeghi et al., 2009; Shirani et al., 2010; Soane and Van Ouwerkerk, 2013).

In mechanized agriculture around the world, soil compaction is known to be a hazardous problem caused by unsuitable farm management. It is estimated that 68.3 million hectares of the worldwide areas is degraded due to compaction (Braunack et al., 2006; Lozano et al., 2013). When soil is compacted, its properties will be negatively affected; for example, there would be poor aeration, low hydraulic conductivity, reduced soil water infilterability, and increased runoff (overland flow). These changes would adversely affect plant root growth. Likewise, the soil compaction destroys soil structure and diminishes soil quality intensifying soil erosion. Both soil pore space and pore size distribution are affected by compressive forces (Håkansson et al., 1998; Hargreaves et al., 2019)

Assessment and quantification of soil compaction is believed to be crucial for soil and water conservation (Boivin et al., 2006). Proctor test is originally proposed by Proctor (1933) for civil engineering, purposed and adapted to be used in agricultural soil mechanics for years. It has several functions in agriculture for determining the soil workability limits, which helps to compare the effect of soil properties and soil management practices on soil mechanical behavior (Zhang et al. 2006; Mosaddeghi et al., 2009; Shirani et al. 2010) and to define a reference bulk density in root growth studies (Reichert et al. 2009; Asgarzadeh et al. 2014; Kodikara et al. 2018). An impact-loading type is used in the test to compact the soil samples with different water contents under a specific compactive effort/energy (CE) in a mold. The compacted dry bulk density is plotted versus water content to determine the maximum dry bulk density (BD_{max}) and optimum water content ($\theta_{optimum}$). The term optimum water content, $\theta_{optimum}$, is replaced with critical water content ($\theta_{critical}$) in agricultural applications to emphasize that while it is an optimum water content for road construction in civil engineering, it will be the worst conditions for tillage and traffic in arable soils (Ekwue and Stone, 1997). The BD_{max} and $\theta_{critical}$ are considered as indices quantifying the soil compactibility (Soane, 1990).

Since soil compaction tests, such as Proctor test, are time-consuming and boring, some scientists have developed PTFs employing conventional regression methods to predict the BD_{max} and $\theta_{critical}$ via easily available properties (Heuscher et al., 2005; Abdelbaki, 2019). Wagner et al. (1994) reported a non-linear relationship between $\theta_{critical}$ and easily available properties (for example, organic carbon, clay, and sand percentage) with high precision. Benites et al. (2007) estimated BD_{max} by regression PTFs (R^2=0.71) using some soil properties, including organic matter, sand, and clay contents as predictors. Over the recent years, new data mining techniques, such as artificial neural network (ANN), have been employed as powerful predictive models to assess the sparsely-available properties in soil science and engineering (Besalatpour et al., 2013; Zolfaghari et al., 2015, Khaboushan et al., 2018; Mohammadi et al., 2020). However, fewer attempts have been made to evaluate the suitability of these techniques for prediction of BD_{max} and $\theta_{critical}$.

The soils in the central and southern Iran are mainly calcareous and gypsiferous, saline, and with low organic matter content (Sarmast et al. 2016). One of the problems in Iranian agriculture is soil compaction. Kerman province, the largest province in Iran, is located in southeast of Iran. This province is one of the most important agricultural regions in Iran, whose main crop is pistachio.

Currently, soil densification has happened in the soils of pistachio gardens since farmers do not frequently apply organic manures and would usually use mineral fertilizers. Therefore, soil has compacted by mechanized cropping practices. Roughly, about 300,000 ha of pistachio gardens are affected by low soil organic matter content and soil densification in Kerman province (Shirani et al., 2010). Therefore, it could be useful to identify the factors leading to soil compaction.

Although there are PTFs available to predict soil compaction parameters in references, most of these PTFs are based on regression and the use of artificial neural network methods and special statistical analysis is very rare, especially in soils of pistachio orchard in the arid regions of Iran. PTFs were mostly site-specific and could not be used for other geographical regions with different climatic and soil attributes. The developed PTFs for temperate and humid regions in the world are obviously not applicable for the aridic soils due to intrinsic differences between the soils of different climatic and/or geographical zones. Clay minerals of the aridic soils are completely different from the temperate soils. Therefore, the current study was aimed to:(i) derive PTFs for estimation of BD_{max} and $\theta_{critical}$ from the easily-available soil properties, such as the sand, silt, clay, gypsum, CCE, and organic matter contents and bulk density, and (ii) use ANN technique in addition to the multivariate linear regression (MLR) method in order to upgrade the PTFs efficiency. The prediction accuracy of both methods was evaluated with well-known statistics.

Materials and Methods

Study region and the soils

The soil samples were collected from different regions of Kerman province, southeast of Iran (Figure 1) (a total area of 186000 km^2, (25^0 55$'$ to 33^0 6$'$ N and 53^0 26$'$ to 59^0 29$'$ E). In each region, the soil samples were taken from the topsoil (0−30 cm) under two kinds of land use, namely planted (pistachio and fruit gardens or crop lands) and unplanted lands (virgin and desert). It was attempted to collect the soil samples with a wide range of intrinsic properties and under diverse land uses which are typical in the region (Table 1). The numbers of sampling from agricultural lands were greater in three cities, namely Kerman, Rafsanjan, and Baft, because other lands in these cities are occupied by urban regions. Moreover, Kerman, Rafsanjan, and Baft are more important agriculture-wise. On top of the pistachio orchards in Kerman, certain plants, such as cereals and vegetables are produced. Rafsanjan is the main center of pistachio production and Baft is a mountainous region producing fruits, like peaches, apricots, cherries, and almonds. Dur to the existence of a high percentage of gravel in the virgin lands, only the agricultural lands in Baft were sampled. This would result in the PTFs which might be applicable for all of the soil types in the region and enhance the reliability of the derived PTFs. In total, 100 soil samples were collected (Table 1). The sampled soil mostly included subgroups of Typic Torrifluvents and Typic Hplocalcids (Soil Survey Staff, 2014). In each sampling location, 6 kilograms of soil were uniformly taken from the topsoil layer and transferred to the laboratory in a plastic bag. After air-drying, the samples were ground and passed through a 2-mm sieve. The sieved samples (< 2 mm) were used for the measurement of soil properties and Proctor compact test.

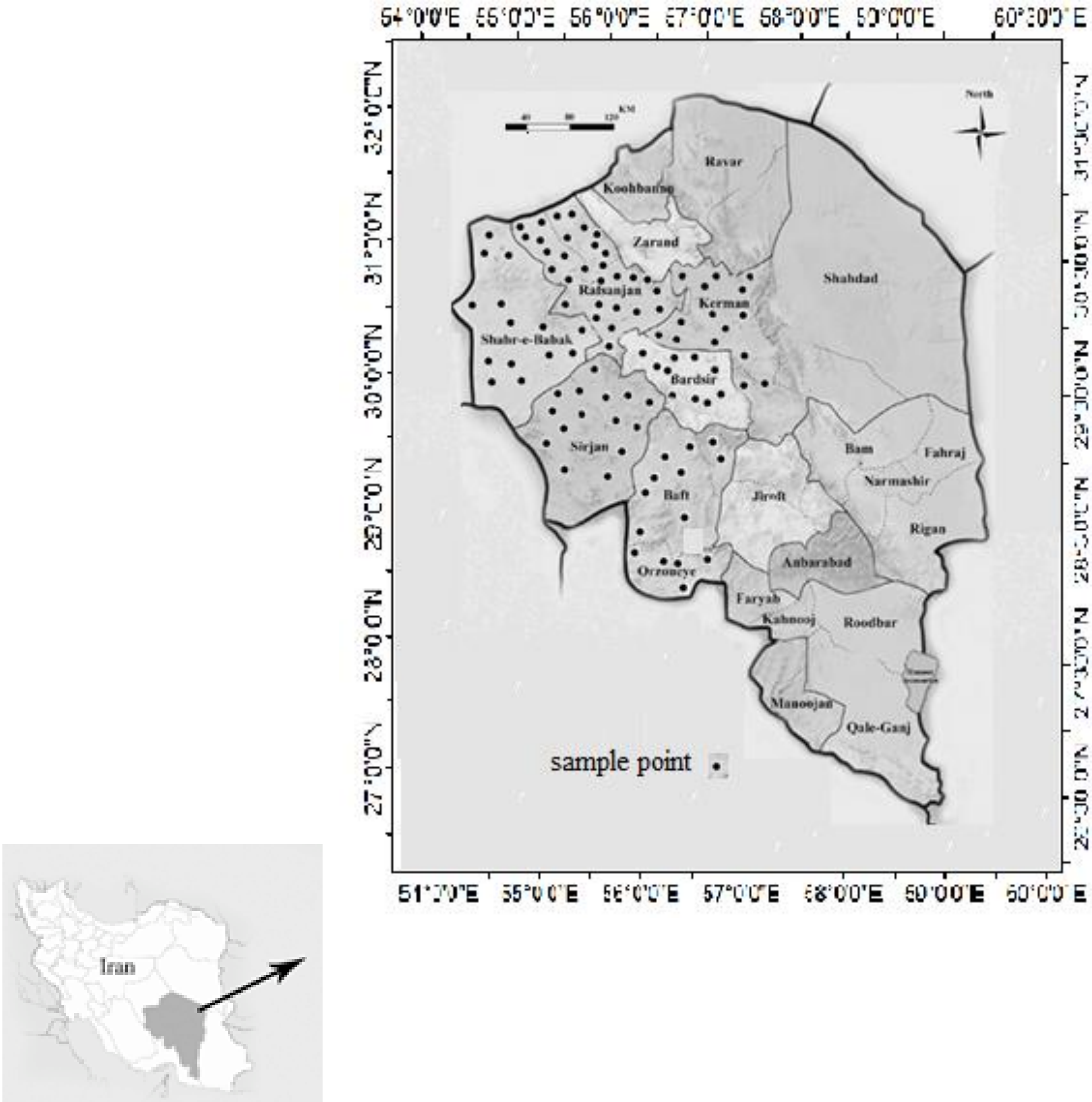

Figure 1. Study area and sampling locations in the investigated area of Kerman province (Iran)

Table 1. The studied regions and number of the collected soil samples for each region and land use

Region/City	Landuse	
	Garden and field	Virgin or desert
Kerman	15	−
Rafsanjan	17	13
Bardsir	7	4
Shahrebabak	6	8
Sirjan	8	7
Baft	15	−
Total	68	32

Measurement of soil properties

A small amount of soil samples was utilized for measurement of intrinsic properties. Soil texture (primary particle size distribution) was determined using the hydrometer method (Gee and Bauder, 1986). The gypsum content was measured via the acetone method (Nelson et al., 1978). The CCE content was determined with the back-titration method (Sims, 1996). Wet-oxidation techniques were applied to measure the soil organic matter (OM) content (Walkley and Black, 1934). Natural bulk density ($BD_{natural}$) was measured employing undisturbed cores with diameter and height of 5 and 8 cm, respectively.

Proctor compaction test

The soil compactibility indices were determined using standard Proctor compaction test. The standard Proctor test, developed for civil engineering applications, employs 25 blows per layer (in total 75 blows) of a 2.5 kg falling hammer from 30.48-cm height (Proctor, 1933; Lambe, 1951). The total height of the compacted soil in the compaction mould was 11.7 cm. Therefore, the thickness of each of the three layers, equally compacted, was about 3.9 cm. The diameter of the compaction mould was 10.16 cm. Thus, a compactive effort (CE) of 600 kN-m/m^3 was produced in this test (ASTM, 1992). A lower energy input (5 blows per layer, in total 15 blows, CE of 120 kN-m/m^3) could be suggested to adapt the test conditions with a lower loading intensity imposed by agricultural machinery tires (Ekwue and Stone, 1997; Barzegar et al., 2000; Shirani et al., 2010). However, the standard Proctor test was employed in this study since it is frequently reported in the literature that BD_{max} of the standard Proctor test is a reference BD in soil and root growth studies (Reichert et al., 2009; Asgarzadeh et al., 2014).

The air-dried soil was prepared and divided into sub-samples with six different water contents. We did our best to select the water contents equally located in the sides of the critical water content ($\theta_{critical}$) (three in the dry limb and three in the wet limb). The soils with different water contents were prepared by water misting and soil remolding to achieve a uniformly-moisturized soil mass. The soil samples were compacted and wet bulk density (BD_{wet}) of the compacted soil was determined. Having the gravimetric water content (θ) of a small sample from the compacted soil was determined and the dry bulk density (BD_{dry}) of the compacted soil was calculated using the following equation:

$$BD_{dry} = \frac{BD_{dry}}{1+\theta}\mathrm{BD} \tag{2}$$

Compaction curves were drawn as BD_{dry} of the compacted soil versus θ. The maximum dry bulk density (BD_{max}) and the corresponding $\theta_{critical}$ were determined based on the compaction curve and were considered as the soil compactibility characteristics.

Derivation of pedotransfer functions and statistical analysis

The PTFs were derived for prediction of BD_{max} and $\theta_{critical}$. Soil texture (primary particle size distribution, sand and clay contents in other words), $CaSO_4$, $CaCO_3$, OM contents, and natural BD were considered as predictors in the PTFs. Two groups of PTFs were developed using MLR and ANNs. The linear regression PTFs and Path analysis were developed to analyze the indirect effects of variables on BD_{max} using SPSS (ver. 16, IBM Com., Chicago, USA).

Path analysis is a statistical technique utilized to examine the comparative strength of direct and indirect relationships among variables. A series of parameters were estimated by solving one or more structural equations in order to test the fit of the correlation matrix between two or more causal models, which were hypothesized by the researcher to fit the data (Lleras, 2005)

The ANN-based PTFs were derived with Matlab (2013), Neural Networks Toolbox (Mathworks, Inc., Natick, MA, USA). A feed-forward propagation technique was applied to derive the ANN-based PTFs. In the ANN, the best model was obtained for BD_{max}, when all the easily-available properties, including sand, silt, clay, OM, $CaSO_4$, and $CaCO_3$ contents, and BD were used in the PTFs. In this method, we used 70% of the data for training, 15% for testing, and 15% for validation.

To assess the model performance, various standard statistical performance evaluation criteria were assessed for the testing data. The statistical measures included were the root mean square error (RMSE), the mean absolute percentage error (MAPE), and coefficient of determination (R^2) between the measured and predicted $\theta_{critical}$ and BD_{max}. The RMSE and MAPE statistics are defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[p(x_i) - M(x_i)]^2} \quad (3)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{[p(x_i) - M(x_i)]}{M(x_i)}\right| \times 100 \quad (4)$$

where $P(x_i)$ denotes the predicted value of observation i, $M(x_i)$ is the measured value of observation i, and n is the total number of observations.

In order to evaluate the importance of input variables in the ANN-based models, the coefficient of determination (R^2) was primarily calculated by importing all of the predictors using the training data. Subsequently, a variable was excluded from the inputs and ANN-based model was similarly developed using the training data and its coefficient of determination was denoted by R^{2*}. Finally, the variable importance (VI) of the excluded variable was calculated via the following equation (Shirani et al., 2015):

$$VI = \frac{R^2 - R^{2*}}{R^2}\mathrm{VI} \quad (5)$$

This equation indicates that summation of all VI values must be equal to 1 and the variable(s) with higher VI value would have greater effects on the dependent variable (output) and its prediction.

Results

Table 2 depicts certain statistical details of the physical and chemical properties of the studied soils. The soil properties have broad ranges with high SD, which are good to develop reliable PTFs for compactibility indices in the region.

Correlation matrix of the measured soil variables is represented in Table 3. Significant positive correlations were observed between BD_{max} and sand content or $BD_{natural}$, but BD_{max} had significant negative correlations with silt, clay, and organic matter contents. The correlations between $\theta_{critical}$ and sand, CCE, and gypsum contents were negative, but those with silt and clay contents were positive. The positive correlation between BD_{max} and $BD_{natural}$ indicated that soils with greater bulk density in the natural condition would have higher maximum bulk density obtained in the Proctor test. The negative correlation between BD_{max} and $\theta_{critical}$ showed that the high maximum bulk density corresponded to the low critical water content, which is in agreement with other studies (Shirani et al., 2010).

Table 2. Statistics of the measured soil properties in the study area

Statistics	Sand	Silt	Clay	OM	$CaSO_4$	$CaCO_3$	BD	BD_{max}	$\theta_{critical}$
	Soil properties								
	kg $100kg^{-1}$						Mg m^{-3}		kg kg^{-1}
Min	44.00	2.64	5.16	0.01	4.19	12.45	1.17	1.53	17.29
Max	86.92	32.25	28.44	0.42	18.47	21.45	1.69	2.10	40.21
CV	0.16	0.36	0.27	0.50	0.23	0.12	0.10	0.07	0.23
Mean	63.40	18.48	18.12	0.20	12.98	16.36	1.54	1.78	28.40
SD	10.19	6.71	4.91	0.10	2.97	1.90	0.16	0.12	6.39

Min= minimum value, Max= maximum value, CV= coefficient of variability, SD= standard deviation.

Table 3. Correlation matrix of the measured soil variables

	Sand	Silt	Clay	OM	$CaSO_4$	$CaCO_3$	$BD_{natural}$	BD_{max}
Silt	-0.890^{**}							
Clay	-0.804^{**}	0.616^{**}						
OM	-0.465^{**}	0.381^{**}	0.454^{**}					
$CaSO_4$	0.153^{ns}	-0.138^{ns}	-0.142^{ns}	-0.257^{*}				
$CaCO_3$	0.197^{ns}	-0.172^{ns}	-0.122^{ns}	-0.252^{*}	0.121^{ns}			
$BD_{natural}$	0.869^{**}	-0.734^{**}	-0.740^{**}	-0.476^{**}	0.094^{ns}	0.119^{ns}		
BD_{max}	0.613^{**}	-0.568^{**}	-0.558^{**}	-0.297^{*}	0.087^{ns}	0.381^{ns}	0.876^{**}	
$\theta_{critical}$	-0.796^{**}	0.490^{**}	0.937^{**}	0.080^{ns}	-0.675^{**}	-0.530^{**}	-0.756^{**}	-0.515^{**}

Ns= non-significant, *: P < 0.05, **: P < 0.01

Linear regression PTFs

The first group of PTFs was derived for soil compactibility indices (BD_{max} and $\theta_{critical}$) using MLR techniques. Tables 4 and 5 illustrate the regression analysis for BD_{max} prediction model and $\theta_{critical}$ prediction model, respectively.

PTFs for the maximum dry bulk density (BD_{max})

Regression analysis (after deleting two outlier observations) showed that $BD_{natural}$, clay, and sand contents could significantly (P<0.001) contribute to the BD_{max} prediction in the PTF . The other variables could not be entered into the PTF (did not significantly affect the BD_{max}) (Table 4). Regression PTF for BD_{max} prediction was derived as following:

$$BD_{max} = 0.603 - 0.0037 Sand\% + 0.003 Clay\% + 0.883 BD_{natural} \quad (6)$$

Variance inflation factor (VIF) implied that the correlation between the entered independent variable was low and the regression model had enough validity (Tables 4 and 5). VIF values lower than 10 implied that the derived model was properly validated (Montgomery et al., 2012; Pal and Bharati, 2019).

The standard regression model for BD_{max} is as following:

$$BD_{max} = -0.339 Sand\% + 0.108 Clay\% + 1.240 BD_{natural} \quad (7)$$

In the standard model, $BD_{natural}$ coefficient is greater than those for sand and clay contents and has the greatest effect on BD_{max} estimation.

Table 4. Regression analysis for BD_{max} prediction model

Predictor	Coef.	SE	*T*	*P*	VIF
Constant	0.603	0.1000	6.00	0.000	
Sand	−0.0037	0.0010	−3.20	0.002	5.3
Clay	0.003	0.0015	1.97	0.049	2.9
$BD_{natural}$	0.883	0.0649	13.59	0.000	4.2

Coef.= Coefficient, SE= Standard error, *T*= *t*-student, *P*= *P*-value, VIF= Variance inflation factor

Table 5. Regression analysis for $\theta_{critical}$ prediction model

Predictor	Coef.	SE	*T*	*P*	VIF
Constant	11.79	2.104	5.6	0.000	
Clay	1.12	0.058	19.1	0.000	1.7
$CaSO_4$	−0.265	0.096	−2.75	0.049	1.7

Coef.= Coefficient, SE= Standard error, *T*= *t*-student, *P*= *P*-value, VIF= Variance inflation factor.

PTFs for critical water content (θcritical)

Regression analysis showed that clay and gypsum contents could be entered into the PTF for $\theta_{critical}$ prediction (Table 5) with significant effects (P<0.05). However, other variables were removed during stepwise regression analysis. The final derived PTF reads as the following:

$$\theta_{critical} = 11.79 + 1.12Clay\% - 0.265CaSO_4\% \quad (8)$$

The standard regression model for $\theta_{critical}$ is as following:

$$\theta_{critical} = 0.858Clay\% - 0.124CaSO_4 \quad (9)$$

The standard model showed that clay content had greater effects on the $\theta_{critical}$ compared with gypsum.

Figure 2 demonstrates path Graph for BD_{max} prediction. As is evident from the standard regression coefficients, $BD_{natural}$ had a great direct effect on BD_{max}. However, sand content had a high indirect effect on BD_{max} via $BD_{natural}$ (1.029), which was as a result, of positive binary correlation between sand content and BD_{max} (Table 6). Meanwhile, its direct effect on BD_{max} was negative. Table 6 shows the direct and indirect effects of sand, clay contents, and $BD_{natural}$ on BD_{max} through Path analysis.

Artificial neural network PTFs for BD_{max} and $\theta_{critical}$

The second group of PTFs was derived for soil compactibility indices (BD_{max} and $\theta_{critical}$) using ANNs. The best neural network topology, obtained by trial and error, was one hidden layer with 10 neurons, LM training algorithm, activity function of *tansig* in the hidden layer and that of *pureline* in the output layer, and a number of training epochs of 100. The R^2 value of the PTFs with all the measured properties are given in Table 7.

The R^2 value of the PTFs for BD_{max} with all of the measured properties (sand, silt, clay, organic matter, gypsum, and CCE contents and natural bulk density) included as predictors was 0.93 for the test data and 0.84 for the train data (Table 7). In order to present a more applicable PTF with a lower number of input variables 3), a PTF was derived using $BD_{natural}$ and sand and clay contents, as predictors with R^2 = 0.81.

For $\theta_{critical}$ prediction, the best model was obtained once all the easily-available properties were used in the PTF. The R^2 of the PTFs with all of the measured properties included as predictors was 0.97 for the test data and 0.91 for the train data (Table 7). A more applicable PTF (with R^2= 0.89) was also derived with three easily-available properties (clay, gypsum, and CCE contents) to predict $\theta_{critical}$. Figures. 3 and 4 exhibit the importance of the coefficients of the input variables to the ANN model for the prediction of BD_{max} and $\theta_{critical}$, respectively. Table 7 represents the results of the comparison of MLR and ANN models using the studied indices.

Table 6. Path analysis of the direct and indirect effects of sand and clay contents and $BD_{natural}$ on BD_{max}. The values on the diagonal indicate direct effects, those on both sides show indirect effects, and column *r* stands for the correlations between these variables and BD_{max}

	Sand	Clay	$BD_{natural}$	*r*
Sand	−0.328	−0.088	1.029	0.613
Clay	0.263	0.108	−0.929	−0.558
$BD_{natural}$	−0.285	−0.082	1.242	0.876

Table 7. Comparison of the performances of the proposed models for the BD_{max} and $\theta_{critical}$ predictions

Parameter		Evaluation criterion		
		MAPE	RMSE	R^2
		MLR model		
$\theta_{critical}$		44.71	24.87	0.51
BD_{max}		16.63	4.56	0.45
		ANN model		
$\theta_{critical}$	test	15.22	6.58	0.91
	train	9.31	2.47	0.97
BD_{max}	test	7.41	2.26	0.84
	train	3.65	0.86	0.93

MAPE: mean absolute percentage error, RMSE: root mean square error, MLR: multivariate linear regression, and ANN: artificial neural network

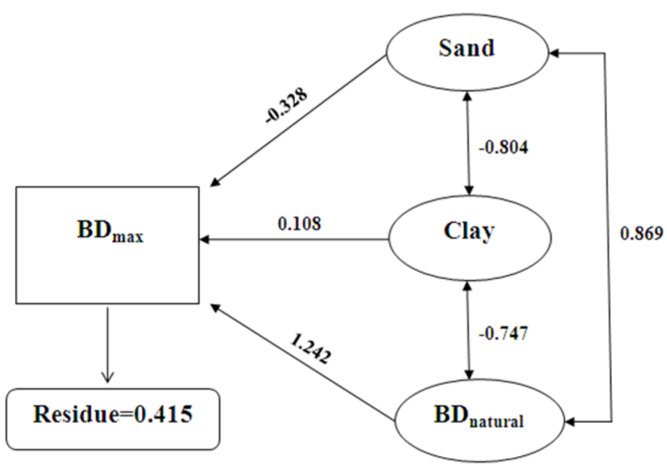

Figure 2. Graph of Path analysis for BD_{max} prediction. One-way arrows and path coefficients show direct effects of sand and clay contents and $BD_{natural}$ on BD_{max}; the two-way arrows show the correlation (*r*) between the variables

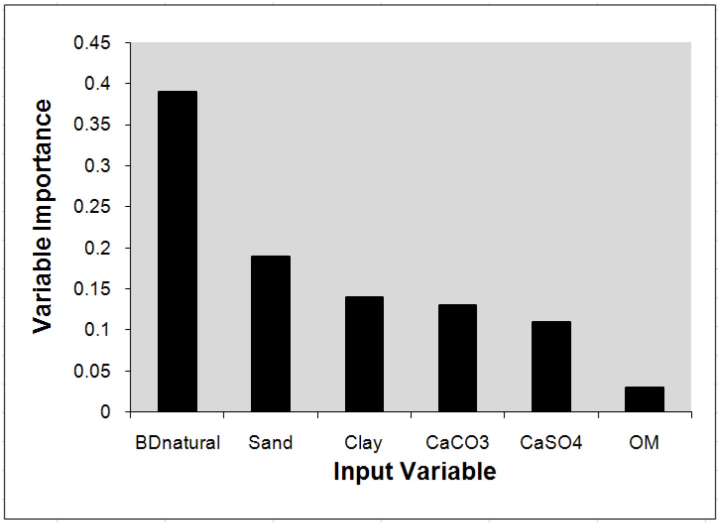

Figure 3. Important coefficients of the input variables to the ANN model for the prediction of BD_{max}

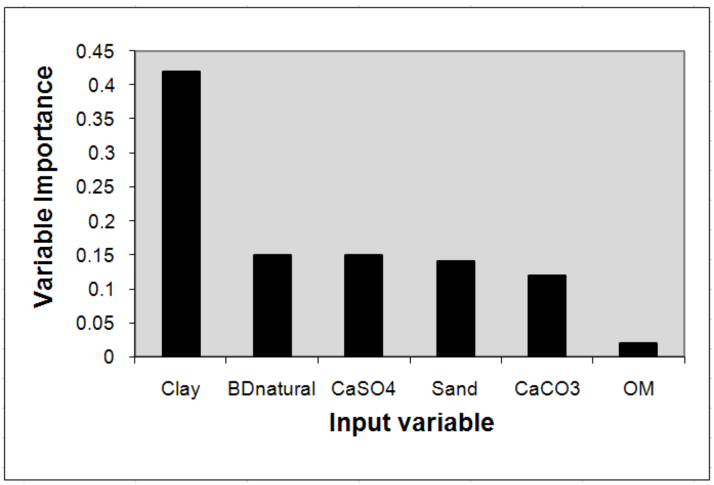

Figure 4. Important coefficients of the input variables to the ANN model for the prediction of $\theta_{critical}$

Discussion

Linear regression

PTFs for BD_{max}

According to the regression models (Eqs. 6 and 7), BD_{max} had a negative relationship with sand content and a positive one with clay content. Nevertheless, the correlation between BD_{max} and sand content was positive and that with clay was negative (Table 3). This discrepancy could be interpreted this way: in the multiple regression analysis, the highest coefficient with positive effect (+1.240) belonged to the $BD_{natural}$; sand and clay contents, with their high correlations with $BD_{natural}$ (Table 3), would have an indirect effect on BD_{max} via $BD_{natural}$. Therefore, Path analysis was done (see subsection 3.1.3) to analyze and interpret the indirect effects of sand and clay contents (via $BD_{natural}$) on BD_{max}.

Analysis of the residuals (errors) showed that the derived PTF (Eq. 6) is a suitable model to predict the BD_{max}. Based on the results using Eq. 7, $BD_{natural}$ had the greatest effect on BD_{max} and then sand and clay contents had, respectively, negative and positive effects on the BD_{max}. Benites et al. (2007) reported that clay, sand, and organic matter contents are highly effective on BD prediction. In this research, OM were not entered into the model since OM values were very low (<0.5%) in the studied soils and were not expected to have significant effects on soil mechanical behavior.

PTFs for $\theta_{critical}$

This finding indicated that clay particles mainly affect water retention in soils with low organic matter content.

Analysis of the residuals (errors) showed that errors are approximately randomly distributed around the reference line similar to BD_{max} and the derived PTF (Eq. 8) is a suitable model to predict the $\theta_{critical}$.

The results from Eq. 8 were not completely consistent with those of Tusheng Ren et al. (2008) who reported a model in which soil texture, bulk density, and organic matter content contributed to estimating $\theta_{critical}$ with $R^2 = 0.92$. (Vaught et al., 2006). They also showed that $\theta_{critical}$ mostly depends on the soil texture with $R^2 = 0.88$. In our study, OM was not entered into the model because its values were very low (<0.5%) and were not expected to have significant effects on soil mechanical behavior.

With regard to $\theta_{critical}$ prediction, clay content was the most effective parameter that could explain the $\theta_{critical}$ variability in the soils of Kerman province. This is in accordance with the reports of other researchers (Vaught et al., 2006). This effect is due to greater water retention of fine (clay) particles resulting from active clay surfaces with high specific area. The increasing effect of the clay content on $\theta_{critical}$ can be observed in Tables 3 and 5, which is in agreement with the results of Vaught et al. (2006) and Lado et al. (2007).

The negative effect of gypsum content on $\theta_{critical}$ (Eq. 8) might be interpreted as the following: gypsum particles do not have high water retention capacity; therefore, existence of this mineral in the soil reduces the soil water content by dilution effect. Additionally, the CCE content had a negative effect on $\theta_{critical}$ (Table 3), but the relation was not as strong as that with gypsum content. This observation may be attributed to low variability of CCE content compared to $CaCO_4$ content in the region (Table 2).

Path analysis

It is principally expected that sand content decrease soil compactibility. However, owing to the positive effect of sand content on $BD_{natural}$, it would indirectly increase BD_{max}. Clay can directly increase the BD_{max} owing to its positive effect on soil compactibility. On the other hand, clay can directly decrease $BD_{natural}$ and has therefore high indirect effects on BD_{max} via $BD_{natural}$ (coefficient 0.929); the binary correlation between clay and BD_{max} is negative (-0.556). afterwards, negative coefficient of sand content and positive coefficient of clay content in the standard regression equation for BD_{max} prediction (Eq. 7) is due to the presence of $BD_{natural}$ in the equation.

When the standard regression equation is derived without $BD_{natural}$ as a predictor, the following equation is obtained:

$$BD_{max} = 0.618Sand\% - 0.058Clay\% \qquad (10)$$

This equation confirms the above-mentioned explanations. It means that with the presence of $BD_{natural}$ in the model (Eq. 7), sand content would have a negative effect on BD_{max} and soil compactibility, and clay content would have negative effect on BD_{max} and soil compactibility, as expected. However, in the absence of $BD_{natural}$, the effect of soil texture on BD_{max} and soil compactibility is completely different (see Eq. 9).

Variable importance in ANN modeling

The results of sensitivity analysis via the relative changes in the coefficient of determination (Eq. 5) showed that the most important variable in BD_{max} prediction using ANNs is $BD_{natural}$, followed by soil textural fractions (sand and clay contents), CCE, and gypsum contents (Fig. 3). In addition to $BD_{natural}$ and soil texture, CCE and gypsum contents were found to be of importance in prediction of BD_{max}. Artificial neural networks are able to derive nonlinear relationships between variables with a high accuracy. This is not possible with binary correlation and linear regression analyses. In other words, on a number of occasions, no linear relations were determined between the variables while ANNs are able to derive strong nonlinear relations between the inputs and outputs. Carbonates and sulfates (CCE and gypsum) are important minerals with high variability in the studied arid soils (Table 2). Therefore, they are expected to significantly affect soil physical and mechanical properties, such as compactibility (Fig. 3). However, OM did not highly contribute to the variability of the BD_{max} (Fig. 3) due to its low values in the region (Table 2).

The results indicated that the highest sensitivity to clay content and the lowest sensitivity to OM content belonged to $\theta_{critical}$ in the studied soils. This finding could be attributed to high water retention of fine particles because of their high surface area. Following the clay content, the effects of gypsum and CCE contents on $\theta_{critical}$ were also high.

Comparison of the models

The comparison of the MLR and ANN models revealed that ANN method provides much more accurate predictions of BD_{max} and $\theta_{critical}$ compared with the MLR model (Table 7). As mentioned above, only the linear effects of the predictors on the target variable can be extracted with MLR method while in several cases, the effects may not be linear in the nature. Meanwhile, neural networks are suitable for modeling nonlinear relationships (Shirani et al., 2015). Nevertheless, the physical effects of the variables in ANN cannot be interpreted via the parameters of the model unlike the regression model, which is one of the main disadvantages of ANN method over regression model (Besalatpour et al., 2013).

Conclusion

1) This study was conducted to derive and evaluate the PTFs with respect to their accuracy and usefulness for prediction of soil compactibility indices (BD_{max} and $\theta_{critical}$) using linear regression and ANN methods in Kerman province, southeast of Iran. Path analysis was also carried out to determine the direct and indirect effects of the predictors.
2) High prediction efficiencies were achieved using the ANNs. However, the accuracy of ANN and MLR methods became almost identical by reducing the number of predictors.

3) Based on the Path analysis, $BD_{natural}$ had a great direct effect on BD_{max} whereas sand had a high indirect effect on BD_{max} via $BD_{natural}$.
4) Sensitivity analysis showed that the most important variable in BD_{max} prediction using ANNs is $BD_{natural}$, followed by sand and clay contents, CCE, and gypsum contents. The $\theta_{critical}$ had the highest sensitivity to clay content and the lowest sensitivity to OM content in the studied soils. This finding could be attributed to high water retention of fine particles because of their high surface area. Following the clay content, the effects of gypsum and CCE contents on $\theta_{critical}$ were also high.
5) The results obtained herein provided information contributing to the prediction of soil compactibility indices and management of arid calcareous soils in Kerman province.

References

Abdelbaki AM. 2018. Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soil. Ain Shams Engineering Journal, 9(4); 1611-1619.

Asgarzadeh H, Mosaddeghi M, Dexter AR, Mahboubi AA, Neyshabouri MR. 2014. Determination of soil available water for plants: consistency between laboratory and field measurements. Geoderma, 226; 8-20.

ASTM, 1992. Standard Test Method for Heat Visible Smoke Release Rates for Materials Products Using an Oxygen Consumption Calorimeter American Society for Testing Materials, West Conshohocken, PA.

Barzegar A, Asoodar M, Ansari M. 2000. Effectiveness of sugarcane residue incorporation at different water contents and the Proctor compaction loads in reducing soil compactibility. Soil Tillage Research, 57; 167-172.

Benites VM, Machado PL, Fidalgo EC, Coelho MR, Madari BE. 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma, 139; 90-97.

Besalatpour A, Ayoubi S, Hajabbasi M, Mosaddeghi M, Schulin R. 2013. Estimating wet soil aggregate stability from easily available properties in a highly mountainous watershed. Catena, 111; 72-79.

Boivin P, Schäffer B, Temgoua E, Gratier M, Steinman G. 2006. Assessment of soil compaction using soil shrinkage modelling: Experimental data and perspectives. Soil Tillage Research, 88; 65-79.

Braunack M, Arvidsson J, Håkansson I. 2006. Effect of harvest traffic position on soil conditions and sugarcane (Saccharum officinarum) response to environmental conditions in Queensland. Australia. Soil Tillage Research, 89; 103-121.

Ekwue E, Stone R. 1997. Density-moisture relations of some Trinidadian soils incorporated with sewage sludge. Journal of agricultural engineering research, 40; 317-323.

Håkansson I, Stenberg M, Rydberg T. 1998. Long-term experiments with different depths of mouldboard ploughing in Sweden. Soil Tillage Research, 46; 209-223.

Hargreaves PR, Baker KL, Graceson A, Bonnett S, Ball BC, Cloy JM. 2019. Soil compaction effects on grassland silage yields and soil structure under different levels of compaction over three years. European Journal of Agronomy, 109; 125916.

Heuscher SA, Brandt CC, Jardine PM. 2005. Using soil physical and chemical properties to estimate bulk density. Soil Science Society of America Journal, 69; 51-56.

Khaboushan EA, Emami H, Mosaddeghi MR, Astaraei AR. 2018. Estimation of unsaturated shear strength parameters using easily-available soil properties. Soil and Tillage Research, 184; 118-127.

Kodikara J, Islam T, Sounthararajah A. 2018. Review of soil compaction: History and recent developments. Transportation Geotechnics, 17; 24-34.

Lado M, Ben-Hur M, Shainberg I. 2007. Clay mineralogy, ionic composition, and pH effects on hydraulic properties of depositional seals. Soil Science Society of America Journal, 71; 314-321.

Lambe T. 1951. Soil testing for engineers. John Willey, New York.

Lleras C. 2005. Path analysis. Encyclopedia of social measurement, 3(1), 25-30
Lozano N, Rolim MM, Oliveira VS, Tavares UE, Pedrosa EMR. 2013. Evaluation of soil compaction by modeling field vehicle traffic with SoilFlex during sugarcane harvest. Soil and Tillage Research, 129; 61-68.
Mohammadi M, Fatemi Aghda SM, Talkhablou T, Cheshomi A. 2020. Prediction of the shear strength parameters from easily-available soil properties by means of multivariate regression and artificial neural network methods. Geomechanics and Geoengineering, 1-13.
Montgomery D, Peck E, Vining G. 2012. Introduction to linear regression analysis. John Wiley & Sons.
Mosaddeghi M, Morshedizad M, Mahboubi A, Dexter A, Schulin R 2009. Laboratory evaluation of a model for soil crumbling for prediction of the optimum soil water content for tillage. Soil Tillage Research, 105; 242-250.
Nelson R, Klameth L, Nettleton W. 1978. Determining Soil Gypsum Content and Expressing Properties of Gypsiferous Soils Soil Science Society of America Journal, 42; 659-661.
Pal M, Bharati P. 2019. Introduction to correlation and linear regression analysis. In Applications of Regression Techniques(pp. 1-18). Springer, Singapore
Proctor R. 1933. Fundamental principles of soil compaction. Engineering News-Record, 111.
Reichert J, Suzuki , Dalvan J, Horn R, Håkansson I. 2009. Reference bulk density and critical degree-of-compactness for no-till crop production in subtropical highly weathered soils. Soil Tillage Research, 102; 242-254.
Sarmast M, Farpoor MH, Boroujeni IE. 2016. Comparing Soil Taxonomy (2014) and updated WRB (2015) for describing calcareous and gypsiferous soils, Central Iran. Catena, 145; 83-91.
Shirani H. 2017. Artificial neural networks with an application approach in agricultural sciences and natural resources, Valiasr University of Rafsanjan Publications. (In persian).
Shirani H, Habibi M, Besalatpour A, Esfandiarpour I. 2015. Determining the features influencing physical quality of calcareous soils in a semiarid region of Iran using a hybrid PSO-DT algorithm. Geoderma, 259; 1-11.
Shirani H, Rizabandi E, Mosaddeghi M, Dashti H. 2010. impact of pistachio residues on compactibility, and permeability for water and air of two aridic soils from southeast of Iran. Arid land research management, 24; 365-384.
Sims J. 1996. Lime requirement. In: Sparks, D.L., A.L. Page, P.A. Helmke, P.N. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnson, M.E. Sumner(Ed.) Methods of Soil Analysis. Chemical Methods. Madison, WI: Soil Sci. Soc. Am. J. Book Series Number 5. ASA and SSSA, pp. 491-515.
Soane B. 1990. The role of organic matter in soil compactibility: a review of some practical aspects. Soil Tillage Research, 16; 179-201.
Soane B, Van Ouwerkerk C. 2013. Soil compaction in Crop production. editors. Developments in agricultural engineering. Amsterdam: Elsevier; p. 662.
Vaught R, Brye KR, Miller D. 2006. Relationships among coefficient of linear extensibility and clay fractions in expansive, stoney soils. Soil Science Society of America Journal, 70; 1983-1990.
Wagner L, Ambe N, Ding D. 1994. Estimating a Proctor density curve from intrinsic soil properties. Transactions of the ASAE, 37; 1121-1125.
Walkley A, Black IA. 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil science, 37; 29-38.
Zhang S, Grip H, Lövdahl L. 2006. Effect of soil compaction on hydraulic properties of two loess soils in China. Soil Tillage Research, 90; 117-125.
Zolfaghari Z, Mosaddeghi M, Ayoubi S. 2015. ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran. Soil use management, 31; 142-154.