International Journal of Forest, Soil and Erosion

ISSN: 2251-824X (E)

Vol 11 No. 2 (2021)

Comparison of regression methods for predicting soil water contents at field capacity and wilting point in Bas Sahara of Algeria

Youcef ABDELHAFID[1&3 *], Mohamed CHEBBAH[2] and Milad Zohra RECHACHI[3]

1Department of Agronomic Sciences, Faculty of Natural Sciences and Life Sciences, University of Biskra, Biskra 07000, Algeria

2 Laboratory of Natural Sciences and Materials, university of Mila, 43000, Algeria

3Center for scientific and technical research on arid regions, CRSTRA, BP N1682, Biskra, Algeria

* Correspondence: abdelhafidyoucef@yahoo.fr

Abstract

In arid regions, the rational management of available scare water ressources depends mainly on soil hydraulic properties (i.e., water retentionand hydraulic conductivity). knowledge of soil water content at field capacity (FC) and permanent wilting point (PWP) are very important parameters in biophysical modelling. However, direct measurement of these parameters are time consuminng and expensive. Using data mining methods enable accurate estimations and good generalisation of these parameters. 120 soils samples were collected from three horizons of soil profils located in Biskra province, bas Sahara of Algeria. The pedologic parameters, such as clay, silt, sand, bulk density, and organic matter content were used as inputs. Three approachs were considered, multiple lineaire regression (MLR), multilayer perception (MLP) and support vector machine (SVM) for predicting soil water contents at field capacity (FC) and permanent wilting point (PWP). The model performance was evaluated and compared with coefficient of deternination (R2), root mean square error (RSME), and mean error (ME) indexes. The results obtained in our study show that both artificial intelligence algorithms MLP and SVM are able to provide better performances than the conventinel MLR. Also, it was found that the MLP model performs somewhat better than SVM in the model prediction stage.

Keywords: Bas Sahara, Field capacity, Multilayer perception, Permanent wilting, point, Regression, Support vector machine

Introduction

Irrigated crops play a vital factor in securing global food production, about 40% of all agricultural production comes from irrigated areas (FAO, 2013). The concept of sustainable water resources management has become an important goal at the basin scale worldwide (karandish and Simunek, 2016). The available water capacity for plants (PAWC) is an important factor for their regenaration, maintenance and productivity. Therfore, this bioavailable water must be taken into account for optimum irrigation scheduling, ecological modeling, and hydrological processes (Debarco et al., 2019). However, in arid and semi-arid regions, it is often observed that water supplies are done in an empirical way and do not take on consideration this data, which results a loss of water, a leaching of nutrients and serious environmental and economic damage (pollution and rising of surface water tables).

The soil hydraulic properties such as soil water contents at field capacity (FC) and permanent wilting point (PWP) are crucial parameters in crop water use efficiency of scarce water ressources (Santra et al., 2018). In lateratuere, it is generally admitted that matric potentials at -33kPa and -1500kPa represent FC and PWP respectively (Botula et al., 2012). However, the direct measurement of these two parameters at multiple locations of soil are generally time consuming, expensive, specially when they are applyed over large area (Haghverdi et al., 2014; McNeill et al., 2018).

To overcome those difficulties, pedotransfer functions (PTFs) are nowadays an alternative approach for estimating them from available or more easily soil informations, such as soil texture data (Cosby et al., 1984; Saxton et al., 1986), organic carbon, and bulk density (Saxton and Rawls, 2006; Rab et al., 2011). Nevertheless, there is no real universal PTFs, the proximity of the study area and the similarity of the parent materials are important elements to assess the potential ability of PTFs (Morvan et al., 2004). However, the most of the research focused on pedotransfer functions (PTFs) were elaborated under humid climatic conditions (Julia-Ferrer et al., 2004). Despite that availability of water is one of the main factors limiting agricultural production in arides and semi-arides regions, little work has been done to develop PTFs for predicting soil water contents (Wösten et al., 2001; Khlosi et al., 2016 ; El Majou et al., 2016 ; Ghorbani et al., 2017; Santra et al., 2018). In Algeria soils, studies describing soil water properties has been the subject of little published works (Dridi and Dilmi, 2011; Dridi and Zemmouri, 2012). As regards the bas Sahara of Algeria, PTFs describing soil water properties over large areas are limited by unavailability of data, but in high demande for sustainable crop production systemes.

In literature, sereral approachs were ofen, considered for developing PTFs such as: multiple linear regression (Gupta and Larson, 1979; Rawls and Brakensiek, 1982; Reichert et al., 2009) and atificiel neurone network (ANN) (Pachepsky et al., 1996; Sharma et al., 2006; Zhang and Schaap, 2017). Recently, support vector machine (SVM) approach was used for estimating soil water retention (Lamorski et al., 2008; Twarakavi et al., 2009; Khlosi et al., 2016; Ghorbani et al., 2017). This technique is now considered as a promissing alternative to ANN approach (Nguyen et al., 2015). For Van Looy et al. (2017) using machine learning algorithms have a mojor advantage that they do not take into account the relations between input and ouput variables. However, using conventional regression approach can be an efficient method if the relationship between dependent and independent variables is not complex. The main objectives of this study were i) to develop three computational intelligence models: multiple linear regression (MLR), multilayer perception (MLP) and support vector machines (SVM) for predicting soil water contents at FC (-33 kPa) and PWP (-1500 kPa). In this work a relatively limited sample data based on soil texture (clay, silt, sand fractions), as well as bulk density and organic matter content were used from the Biskra province, south-east of Algeria and ii) to evaluate their performances for the prediction of FC and PWP in order to develop point PTFs. In our knowledge, using machine learning techniques for predicting soil water content is the first in our region.

Materials and methods

Study area and data

As shown in Figure 1, Biskra province (Ziban region, eastern Algeria) is located at the foot of the saharan Atlas and covers about 21 509.80 km2. It is known for its phoenicicol heritage estimated at 5 million of date palms and 98.478 ha of irrigated area (Agence National d'Amenagement du Territoire (ANAT), 2003). This agro-ecological zone belongs to the Saharan bioclimatic stage. The mean annual rainfall is low, rarely exceeding 145 mm with very limited number of rainy days per year (25 days.year-1).The mean annual temparature is around 22° C with very high amplitudes going up to 50°C. The mean humidity is 43% and the annuel evapotranspiration is 1889 mm (Dubost and Larbi-Youcef, 1998). The four major soil types in this region, according to Word Reference Base classification (IUSS Working Group, 2014) are Fluvisols, Calcisols, Gypsisols, and solonshaks Figure 2 and Table 1. Fuvisols are mainly located along the main wadi, the halomorphic character appeared in depth near the saline soils. Soils under Gypsisols are mainly occupy the Quaternary formations along the Atlas range. They join forces in the central part of the region with gypsum accumulations that are increasing in size towards the south. They are described most often in the oasis of Tolga, Lahdjeb and Lioua. Solonshaks occupy mainly the low-lying areas of the eastern region near the Chott Melghir, north of Oumache palm groves and Outaya plain.

ISSN: 2251-824X (E)

Vol 11 No. 2 (2021)

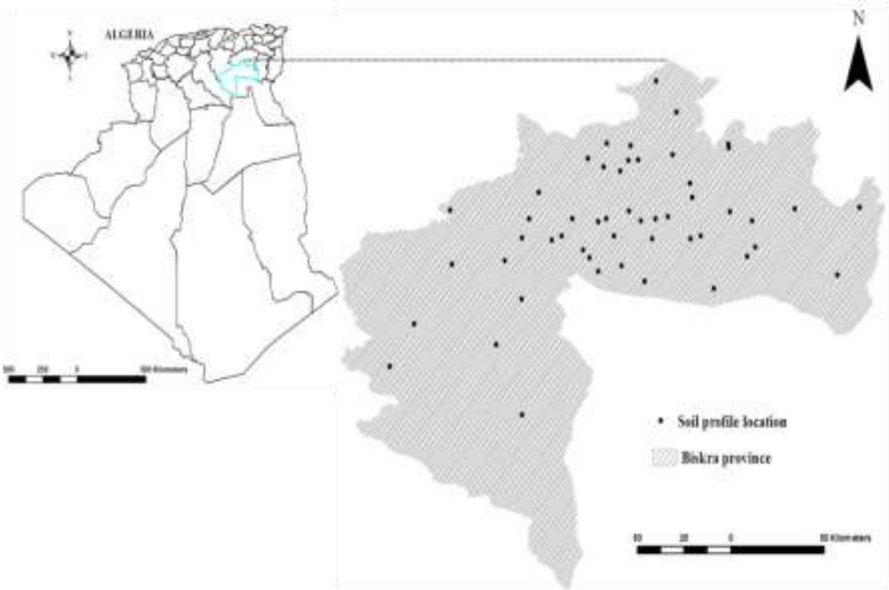

Figure 1. Location of the study area with soil profile location

A total of 120 soil samples were collected from various localities in this area. All collected soil samples were analysed in the laboratory. The bulk density was determined by soil cylinder cores (100 cm3) method, the particle-size analysis (clay, silt and sand contents) were determined by the Robinson's pipette method. Organic matter content (OM%) was determined by Walkley and Black method, soil water content at -333 and 15000 bars were determined by pressure plate apparatus method.

Table 1. Spatial distribution of WRB soil groups in Biskra Province

Reference soil groups	Reference soil groups with principal qualifiers	Soil profiles	Samples
Fluvisols	salic Fluvisols	5	15
	Haplic Fluvisols	3	9
Gypsisols	Haplic Gypsisols	9	26
	Calcic Petric gypsisols	1	3
Solonchaks	Gypsic Solonchaks	4	14
	Salic solonchak	6	21
	Solonchak-solonetz	10	31
	Gleyic Solonchaks	1	3

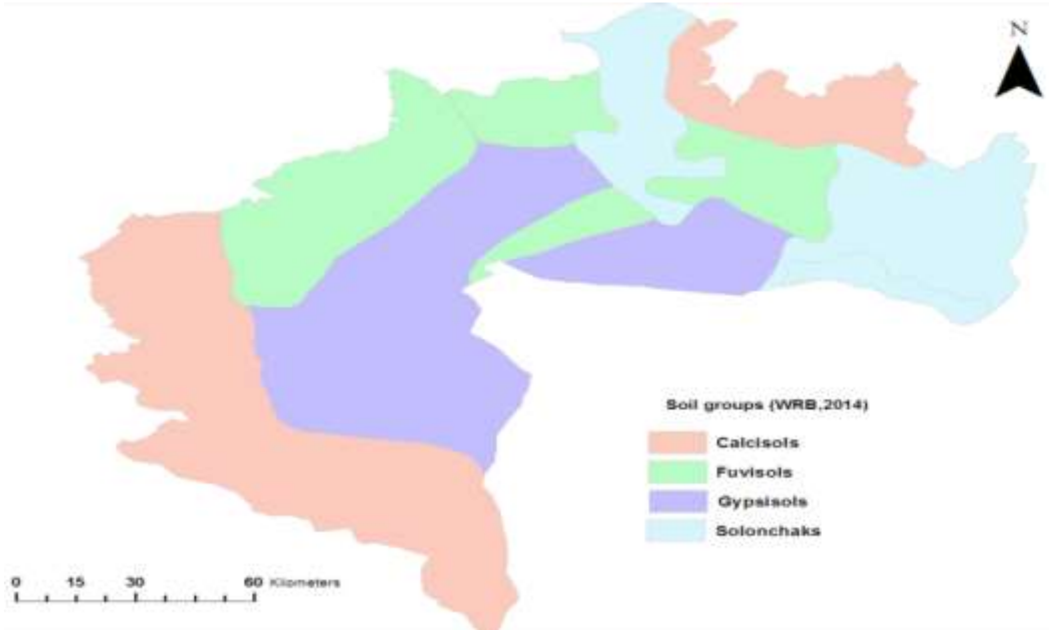

Figure 2. Spatial distribution of WRB soil groups in Biskra Province

Statistical modeling

In the present study, the statistical modeling was done in Stastica (Stasoft 2010). The data were normalised, and randomly divided into datasets, training and testing in a ratio of 70% and 30%, respectively.

The particle size distribution (SD), bulk density (BD), and organic matter content (OM), were used as predictors for predicting FC and PWP. Three modeling approachs were selected: multiple linear regression (MLR), multilayer perception (MLP) and support vector machines (SVM). The multiple linear regression approach was widely used in soil sciences for developing PTFs (Khlosi et al., 2016). The input variables were selected by backward stepwise method (Merdum et al., 2006; Rab et al., 2011). The PTF were developed using a equation of the form:

$$Y = b_0 + b_1X_1 \dots + b_5X_5 + b_6X_1^2 + \cdots b_7X_5^2 + b_8 X_1X_2 + \cdots + b_nX_4X_5 \quad (1)$$

Where Y is the dependent variable, b_0 is the intercept, $b_{01} \dots b_n$ are regression coefficients and $X_1 - X_5$ are independent variables.

The second approach used in this study was the multilayer perception neural network (MLP), which is a feed forward backpropagation articitial neurone network (Pachepsky et al., 1996; Minasny and McBratney, 2002). This network contains input, hidden, and output layers (Figure 3). The hidden layer $y_k(k = 1 \dots K)$ extract useful information from the input layer $x_j(j = 1 \dots J)$ for predicting the target (output layer) (Zhang and Schaap, 2017).

The input vector of neurons is weighted, summed, and biased to produce the hidden neurons

$$y_k = \sum_{j=1}^{J} W_{jk}X_j + b_k, \tag{2}$$

Where J and K are the number of neurons in input and hidden layers. (W_{jk}) is weighted a bias(b_k). The hidden neurons y_k are then operated by an activation or transfer function (f) to produce

$$r_k = f(y_k) \tag{3}$$

The activation function reflet the nonlinearity in the input-ouput relationship. The ouput from the hidden layer is subjected to the similar procedure in equation (2) as follows :

$$v_l = \sum_{K=1}^{K} u_{kl} r_k + b_l, \tag{4}$$

And transformed by another activation function (f) to produce the ouput z :

$$z_l = f(v_{kl}) \tag{5}$$

The weights and biases were obtened in ANN by minimizing the following objective function through an iterative procedure,

$$O(w_{jk}, b_k, u_{kl}, b_l) = \sum_{n=1}^{N_s} \sum_{m=1}^{N_p} \left[t_{n,m} - t'_{n,m}(w_{jk}, b_k, u_{kl}, b_l)\right]^2 \tag{6}$$

Where N_s is the number of samples, N_p the number of parametrs, and t and t' are mesured and estimated variables.

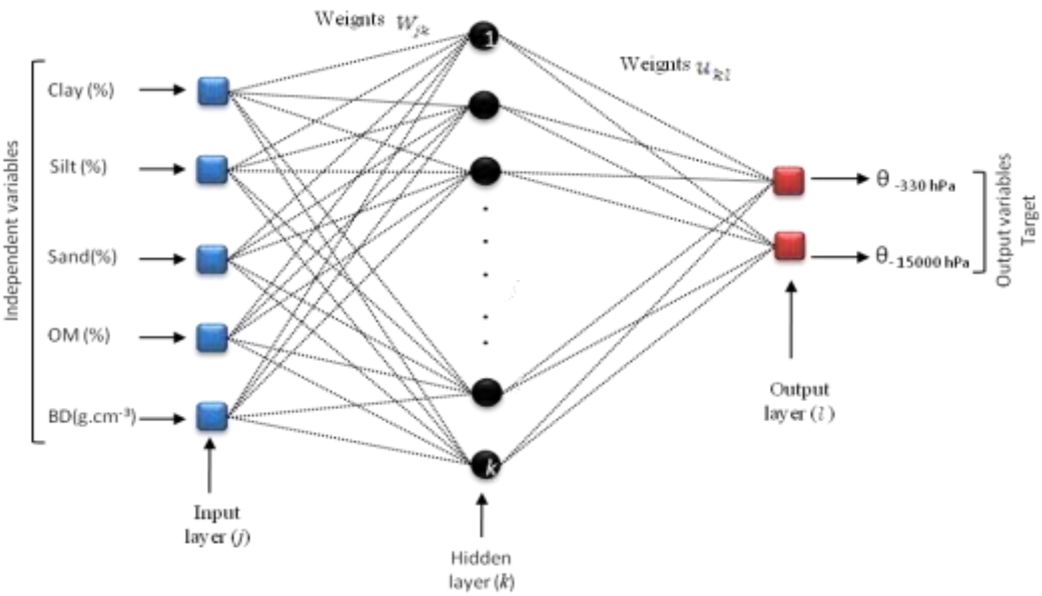

Figure 3. MLP model used in this study

ISSN: 2251-824X (E)
Vol 11 No. 2 (2021)

The third approach called support vector machines (SVM), introduction by Vapnik (1995) (Van Looy et al., 2017). The architecture of a SVM is similar to that of a MLP, but the training algorithm is significantly different (Skalova et al., 2011). The support vector regression (SVR) is used in literature to describe regression with SVM (Karandish and Simunek, 2016). The objectif of SVR modelling is to estimate a function according to a given data set, one output variable based on a set, $\{(x_i, y_i)\}_n$, where x_i denotes the input vector refer to independent variables, y_i de notes the output value refer to soil water contents, and n is the number of samples ($n = 120$). The regression function in SVM approach uses the following formula :

$$f(x) = \sum_{i=1}^{n} \omega \Phi(x) + b \tag{7}$$

Where Φ (x) de notes nonlinear mapping that map data into better representation space. ω_i and b are weights and a bias term estimated by minimizing the following structural risk function:

$$R = \frac{1}{2}|\omega|^2 + C\sum_{i=1}^{n} L_\varepsilon(f\xi + \xi^*) \tag{8}$$

Where R is the sum of the empirical risks and a complexity term. $|\omega|^2 = \sum_{i=1}^{l} w_i^2$. ξ, ξ^* are slak variables, $C > 0$ is the cost parameter, n is the sample size,and ε is the insensitive loss function(Twarakavi et al. 2009).

Subject to

$$\begin{cases} y_i - f(x) - b \le \varepsilon + \xi_i^* \\ f(x) + b - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0, \; i = 1, 2 \ldots\ldots n \end{cases} \tag{9}$$

In SVM models, the radial basis function (RBF) with parameters (C, ε, γ) was used as the kernel function for FC and PWP modelling. This function has the following form :

$$k(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \tag{10}$$

Where γ is the kernel parameter. The different combinison of C,ε,γ parameters determine the accuracy of the SVM model.

In this study, three statistical indices were selected to evalute the model performances: (1) coefficient of correlation (R2), wich describes the proportion of the total statistical variance in the observed dataset that be explained by the model ; (2) the root-mean-square of prediction error (RMSE), a mesure of the overall performance across the entire range of the dataset and (3) the mean of prediction (ME) wich indicates whether there is a systematic errors between measurements and model estimation. Negative ME values indicate an average underestimation of the quantity being evaluated, while positive values indicate an overestimation of target variables. These statistical indices are the most common metrics used to evalute models performance (Van Looy et al., 2017).

ISSN: 2251-824X (E)
Vol 11 No. 2 (2021)

$$R^2 = 1 - \frac{\sum_1^N (Y_i - \hat{Y}_i)^2}{\sum_1^N (Y_i - \overline{Y}_i)^2} \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2} \quad (12)$$

$$EM = \frac{1}{N} \sum_1^N (Y_i - \hat{Y}_i) \quad (13)$$

Where Y_i denotes the mesured value, $\hat{Y}_i$ is the estimated value, $\overline{Y}_i$ is the average of the measured value, and N is the number of samples.

Results

Descriptive statistics of soil data for training and testing data sets are given in Table 2. The results show, both data sets illustrated relatively similar statistical characteristics, for intance, the clay, sand and silt contents, and bulk density were ranged between 2-68%, 2-94%, 3-85%, and 1.06-1.94g.cm-3, respectively, for training data set. FC and PWP varied from 0.03 to 0.46 g/g and from 0.016 to 0.36g/g, respectively. Clearly, the presence of different types of soil in Biskra province cause such large rangs in physical properties.

Table 2. Descriptive staistics of soil data for the training and testing sets

Variable	Training data set				Testing data set			
	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD*
Clay, %	0.5	68	31.1	0.177	03	53	26.47	0.162
Silt, %	01	81	38.3	0.160	06	65	36.52	0.144
Sand, %	02	96.6	30.6	0.272	02	91	35.44	0.264
BD, g/cm^3	1.028	1.94	1.451	0.183	1.058	1.77	1.418	0.157
OM, %	0.1	2.86	0.994	0.008	0.1	03.18	1.14	0.008

CEC (cmole.kg^{-1})	2.40	48	23.16	15.909	3.2	48	22.61	14.845
EC (dS/m)	0.60	148.2	20.41	24.216	0.4	96.50	14.61	20.02
pH	6.86	8.80	7.62	0.406	6.99	8.5	7.75	0.367
$CaCO_3$(%)	1.7	63.80	35.22	0.148	1.5	53.9	32.04	0.119
Gypsum(%)	01	62.08	12.52	0.147	01	64	9.43	0.173
FC, g/g	5.67	57.63	34.82	0.111	9.55	55.98	35.47	0.095
PWP, g/g	2.58	51.60	24.72	0.103	6.53	53.22	27.33	0.105

*S.D Standard deviation

The results of the correlation coefficients between soil data given in Table 3, show that all input variables were statistically significant. Both FC and PWP had a strong correlation ($r > 0.70$) with clay and sand. BD, silt, and OM had a relatively loose relationship with both FC and PWP. Furthermore, sand and BD have an inverse relationship with FC and PWP. These results agreed with those of Botula et al. (2012), Ostavari et al. (2015) and Ghorbani et al. (2017).

Table 3. Coefficients of correlation between soil attributes and soil water contents

Variable	Clay	Silt	Sand	BD	OM	FC	PWP
Clay, %	1						
Silt, %	0.225	1					
Sand, %	-0.744***	-0.737***	1				
BD, g/cm^3	-0.244	-0.333*	0.351*	1			
OM, %	0.374***	0.325*	-0.450***	-0.434***	1		
FC, g/g	0.831***	0.462***	-0.796**	-0.446***	0.489***	1	
PWP, g/g	0.781***	0.414***	-0.724***	-0.454***	0.492***	0.922***	1

* Correlation is significant at the 0.01 level; ** Correlation is significant at the 0.001 level ; *** Correlation is significant at the 0.0001 level

The results obtained in Table 4, confirm the above findings in Table 3. According to R2, RMSE and ME indexes, the 2FC and 2PWP models appear to be better in MLR approach during the training and testing data sets. The predictor variables explained 78% and 73% in total variation at FC and PWP respectivly.

Table 4. Statistics performances of the MLR models

Model	Inputs	Equation	Training			Testing		
			R^2	RMSE	ME	R^2	RMSE	ME
1FC	PSD	0.112 + 0.663(C) + 0.303(Si)	0.770	0.040	0.000	0.791	0.035	0.003

2FC	PSD, BD	0.838 - 0.663(Sa) - 0.303(BD) + 0.631(BD*Sa) + 0.945(BD*C)	0.788	0.038	0.000	0.821	0.035	0.003
3FC	PSD, BD, OM	0.838 - 0.663(Sa) - 0.303(BD) + 0.631(BD*Sa) + 0.945(BD*C)	**0.788**	**0.038**	**0.000**	**0.821**	**0.035**	**0.003**
1PWP	PSD	0.116 + 0.620 (C^2) + 0.378(Si)	0.667	0.044	-0.007	0.647	0.040	-0.000
2PWP	PSD, BD	0.306 + 0.290(Si) - 0.321(BD) + 0.592(Si)	**0.727**	**0.040**	**-0.007**	0.699	**0.037**	-0.001
3PWP	PSD, BD, OM	0.580 - 0.236(Sa) - 0.344(BD) + 0.945(C^2)	0.715	0.041	-0.007	**0.699**	0.040	**-0.000**

BD in conjonction with clay and sand, appears to improve the model performance for predicting FC and PWP, while the influence of OM contents was insignificant. The performance indices R^2 RMSE and ME were also, better at FC than at PWP. Both models included sand, bulk density, BD-clay, BD-sand for FC and sand, BD, and quadratic clay for PWP. This result is in agreement with those of Rab et al. (2011), and Santra et al. (2018).

The Table 5 shows the statistic performances of the models in MLP approach during training and testing data sets. They reveal that the 3FC, and 2PWP models were better as indicated by R2, RMSE and EM values. The optimal model obtened has three layers with single hidden layer (Figure 3).

Table 5. Statistic performances of the MLP models

Model	Inputs	Architecture	Training			Testing		
			R^2	RMSE	ME	R^2	RMSE	ME
1 FC	PSD	3 - 9 -2 (BFGS)	0.826	0.039	-0.000	0.855	0.036	0.01
2 FC	PSD, BD	4 - 3 -2 (BFGS)	0.834	0.034	**0.000**	0.855	0.043	0.004
3 FC	PSD, BD, OM	5 - 5 -2 (BFGS)	**0.840**	**0.033**	-0.001	**0.869**	**0.028**	**0.001**
1 PWP	PSD	3 - 9 -2 (BFGS)	0.730	0.034	-0.000	0.742	**0.029**	0.004
2PWP	PSD, BD	4 - 3 -2 (BFGS)	**0.787**	**0.035**	**-0.000**	**0.778**	0.046	0.007
3 PWP	PSD, BD, OM	5 -5 - 2 (BFGS)	0.785	0.035	-0.000	0.756	0.034	**0.004**

The Table 6 indicates the statistic performances of the models in SVM approach. According to R2, RMSE and EM values, both models 3FC and 2PWP performed well during training and testing period. The performance indices (R^2 RMSE and ME) were also better at FC than PWP.

Table 6. Statistic performances of the SVM models

Model	Inputs	Parameter (C, γ, ε)	Training R^2	Training RMSE	Training ME	Testing R^2	Testing RMSE	Testing ME
1FC	PSD	20, 2.9, 0.85	**0.844**	0.038	-0.005	**0.845**	0.045	0.005
2FC	PSD, BD	10, 2.1, 0.5	**0.855**	**0.032**	-0.001	0.845	0.042	0.005
3FC	PSD, BD, OM	40, 0.25, 0.62	**0.850**	0.034	**-0.000**	**0.851**	**0.029**	**-0.001**
1 PWP	PSD	30, 1.75, 0.75	**0.722**	0.043	-0.003	**0.740**	0.050	-0.006
2PWP	PSD, BD	30, 1, 0.8	**0.774**	**0.038**	-0.011	**0.776**	0.048	-0.003
3PWP	PSD, BD, OM	30, 0.29, 1	0.760	0.038	**-0.007**	0.756	0**.034**	**-0.002**

Fig. 5 shows the comparative plots between the results modeled by the SVM pproach and those measured values for testing data set, which the correlation coefficients are all above 0.80, this demonstrates that SVM technique is capable for predicting FC and PWP.

Discussion

Selecting input variables is an important step for developing prediction models (Ghorbani et al., 2017). Using the texture data, organic carbon, and bulk density as predictor variables are widely taking on consideration for developing PTFs. Additional parameters such as soil particle size and distribution indices, pH, and terrains attributes are rarely used (Wôsten et al., 2001; Reichert et al., 2009; McNeill et al., 2018). Pollaco (2014) provides a simple model for estimating FC and PWP from soil texture and bulk density, while Ghorbani et al. (2017) evaluated the performance of three approachs for estimating FC and PWP using soil texture, bulk density and organic matter content.

In other ways, the size of data can, also determine the predictibility of PTFs. Developing PTFs although less 200 datasets were generaly reported in different studies (Sharma et al., 2006; Santra et al., 2018; Li et al., 2019). For Santra and Das (2008) data with about 100-200 samples may be sufficient to develop reasonable PTFs.

From the Table 3, clay and sand data have a strong relationship with soil water contents. Both FC and PWP were positively correlated with clay contents, silt and OM content, and negatively correlated with sand content, and BD. Our finding agreed with those found by Ostavari et al. (2015), Ghorbani et al. (2017) and Li et al. (2019). As predector variable, clay increases the specific surface area of the soil matrix and favors the occurrence of micropores that generate capillarity (Hillel 1998). Increasing BD decreases porosity and thus soil water retention capacity.

The results obtained in Table 4, confirm the above findings in Table 3. According to R2, RMSE and ME indexes, the 2FC and 2PWP models appear to be better in MLR approach during the training and testing data sets. The performance criteria were also bette rat FC than at PWP. Correlation coefficients are higher at FC than at PWP. The resultats are consistent with those reported in other arid soils (Santra et al., 2018 ; Qiao et al., 2018). OM contents have ofen included as an input variables for prediction FC and PWP (Cornelis et al., 2001; Botula et al., 2012; Nguyen et al., 2015), but, in this study, OM contents did not improve the prediction of FC and PWP. The possible reason may be due to their smaller values and to a higher variation of clay contents (or sand content). The results are similar to those obtained by Minasny and McBratney (2002), Rab et al. (2011), Khlosi et al. (2016), and Santra et al. (2018), who also found soil water content variations were not affected by lower OM contents. Including BD in the models appear to improve soil water retention at FC. In addition, BD plays an important role on the retention water capacity at this level of potential. Increasing BD resultant in the destruction of the pedostructure and pore architecture, leading to reduce the available volume for soil water storage (Hillel, 1974). Furthermore, sand had an inverse influence on soil water contents in MLR approach. Increasing sand particle fraction in soil, loading to increase soil macropores which results a decline in water retention.

The Tables 4-6 given the detail of the model performances in MLR, MLP and SVM approach for predicting FC and PWP. According to R2, RMSE both artificial intelligence algorithms MLP and SVM perform better than the conventional method MLR.

Therfore, the ME values were smaller in MLP approach than those in SVM and MLR approach, which indicates that the quality of estimation in MLP model is better than in SVM and MLR models. For Niu et al. (2019), the MLR approach can only handle conventional linearity, rather than the inherent nonliear relationship in this cas, leading to lower generation capability, whereas, both MLP and SVM approach were able to get the utmost out of the mapping function to map the training data into the high-dimensional feature space, and the carefully choose the appropriate optimization stratigies to find the solution that minimizes the total training error. Moreover, as compared with SVM approach, the models developed in MLP approach had the smallest value of the RMSE as well as higher value of R2 , but the difference in performance with t-statistique ($P < 0.05$) was non significant. These results are comparable to those of Skalova et al. (2011) who also reported that the ANN models performed slighly better than MLR and SVM models with limeted data, and they hypothesized that it is advisable to use combined SVM/MLP models.

However, others studies have reported that the SVM gave the best performance compared by ANN (Twarakavi et al., 2009). For Lamorski et al. (2008) and Khlosi et al. (2016) the SVM models perfomed better than MLP and MLR, but at some matric potentials. However, there is not an universally applicable conclusion, because many studies have also reported that the influence of the number of data or other statistical data set properties can influence the model performances.

The RMSE values for our SVM model are smaller than those obtained by Twarakavi et al. (2009), Svalova et al. (2011), and khlosi et al. (2016).

As seen in Table 7, the RMSE values in MLP approach were smaller as compared to those with SMV and MLR approach during the testing phase. They varied from 0.002 - 0.048 for SVM models and varied from 0.031 to 0.057 and from 0.035 to 0.040, respectively for MLP and in MLP models, the R2 values varied from 0.739 to 0.879 and show more

ISSN: 2251-824X (E)
Vol 11 No. 2 (2021)

accurate prediction than with SVM and MLR models, for wich the values varied from 0.647 to 0.821 and 0.692 to 0.849, respectively.

In Overall, the performance of all models developed during the testing period is showed in Fig. 4, the 3MLP model was selected as the best-fit model for predicting FC 2MLP model for predicting PWP. Evalution indices were also better at field capacity than wilting point for all models. Ghorbani et al. (2017) are found that both MLP and SVM models were classified as good performance. Haghverdi et al. (2012) showed that the performance of either of data mining methods depend on the PTF type, and database characteristics.

Finally, the MLP models with sand, clay, silt, bulk density as predictors appear the best for predicting soil water content at field capacity and wilting point.

Table 7. Goodness-of-fit for MLR, MLP and SVM models in predicting soil water contents at FC and PWP

Variable	Input variable	MLR			MLP			SVM		
		R^2	RMSE	ME	R^2	MSE	ME	R^2	RMSE	ME
1FC	PSD, BD	0.821	0.035	0.003	0.840	0.033	-0.001	0.851	0.029	-0.001
2PWP	PSD, BD	0.699	0.037	-0.001	0.778	0.046	0.007	0.776	0.048	-0.003

International Journal of Forest, Soil and Erosion
ISSN: 2251-824X (E)
Vol 11 No. 2 (2021)

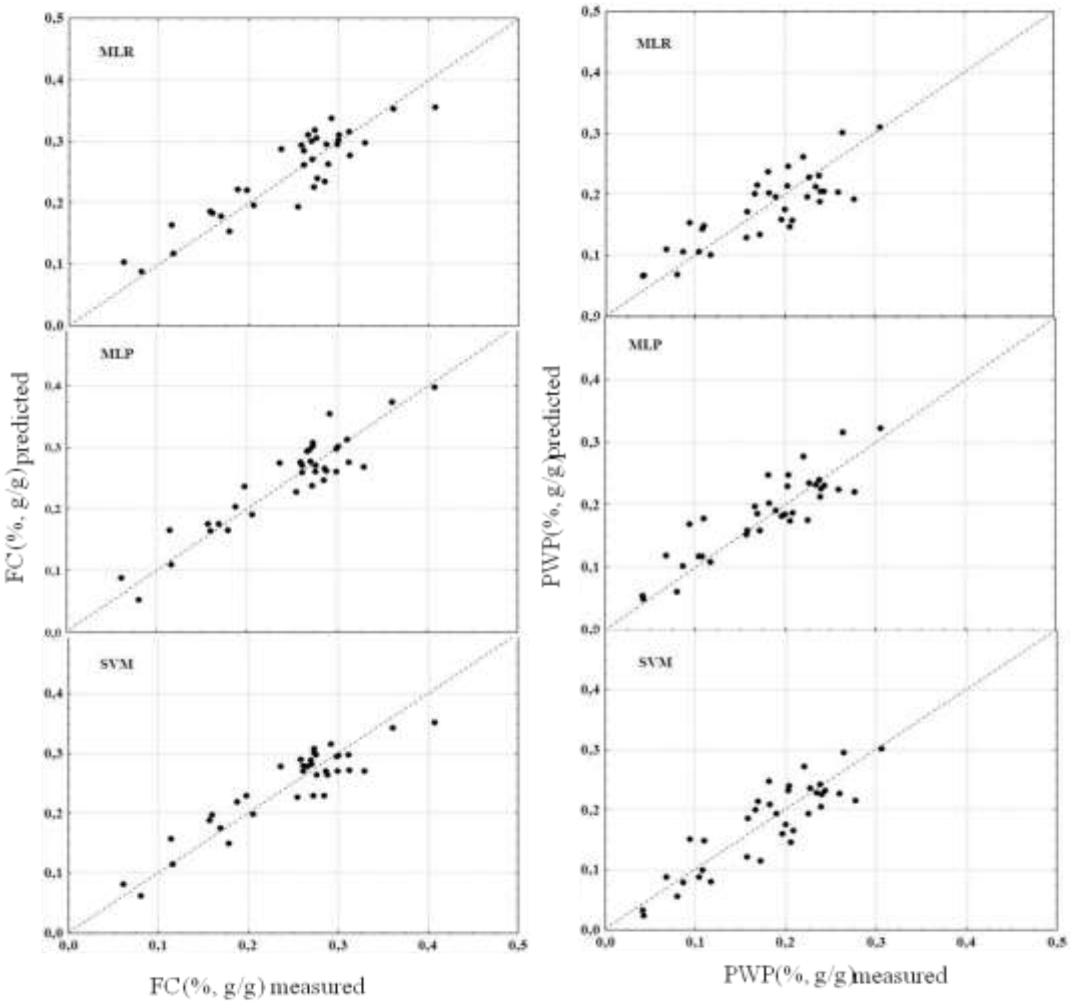

Figure 4. Scatter plots of measured and predicted FC and PWP using MLR, MLP and SVM models in testing set.

Conclusions

In this study we investigate the performance of three regression approach : MLP and SVM as artificial intelligence algorithms and MLR conventional method for estimating the soil water contents at field capacity (FC) and point wilting permanent (PWP). The predictive capabilities of the trained model was accessed using some evaluation criteria. To achieve this objective, relative data samples from Biskra province in bas Sahara of Algeria were employed. Three input variables including clay, silt and sand content, bulk density and organic matter content. The obtened results indicate that the models elaborated by artificial intelligence algorithms (MLP and SVM) provide better prediction performance than conventional method (MLR). That is indicate that the artificial intelligence approach were powerful tools in modeling soil water contents. As next step, comparing the results of the models, it twas seen that overall performance of MLP model was better than that of SVM, but this difference was statistically not significant. Therfore, the results of this study were highly encouraging and suggested that SVM and MLP methods were promissing in modeling PTF for estimating soil water contents in our region.

Acknowledgements The authors wish to express their gratitude to National Hydraulic Ressources Agency (ANRH) and Agence National d'Amenagement du Territoire (ANAT) for providing us the documents related to the study area.

References

ANAT (Agence National d'Amenagement du Territoire) (2003) Schéma directeur des ressources en eau de la wilaya de Biskra. Dossier agro-pédologique, 175 p.

Botula, Y-D., Cornelis, W.M., Baert, G., Van Ranst, E. (2012) Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R.Congo). Agric Water Manag 111:1-10. DOI: https://dx.doi.org/10.1016/j.agwat.2012.04.006

Cornelis, W.M., Ronsyn, J., van Meirvenne, M., Hartmann, R. (2001) Evaluation of pedotransfer functions for predicting the soil moisture retention curve. Soil Sci Soc Am J, 65, 638-648.

Cosby, B.J., Homberger, G.M., Clapp, R.B., Ginn, T.R. (1984) A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour Res, 20(6), 682-690.

Dobarco, M.R., Cousin, I., Le Bas, C., Martin, M.P. (2018) Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma, 336, 81-95. DOI: https://doi.org/10.1016/j.geoderma.2018.08.022

Dridi, B., Dilmi, A. (2011) Poids des différentes caractéristiques des sols dans l'estimation de leur rétention en eau, cas des sols d'Algérie. Étude et Gestion des Sols,18(4),247-258.
Dridi, B., Zemmouri, S. (2012) Fonction de pédotransfert pour les sols de la plaine de la Mitidja (Algérie): recherche de paramètres les plus pertinents pour la rétention en eau. Biotechnol Agron Soc Environ,16 (2), 193-201.

Dubost, D., Larbi youcef, Y. (1998) Mutation agricoles dans les oasis algériennes: l'exemple des Ziban. Sécheresse, 9, 103-110.

El Madjou, H., Kaba, R., Almesber, W., Bruand, A. (2016) Validation de l'estimation des propriétes de rétention en eau de sols syriens à partir de fonctions et classes de pédotransfert développées pour des sols français. Étude et Gestion des Sols, 23, 112-123.

F.A.O (Food and Agriculture Organization of the United Nations) (2013) Climate change, water and food security, 36, Rome, Italy.

Ghorbani, MA., Shamshirband, S., haghi, D.Z., Azani, ABonakdari H., Ebtehaj, I. (2017) Application of firefly algorith-based support vector maxchines for prediction of fiel capacity and permanent wilting point. Soil Tillage Res,172, 32-38. DOI: https://dx.doi.org/10.1016/j.still

Gupta, S.C., Larson, W.E. (1979) Estimating soil water retention characteristics from particle size distribution. Organic matter percent and bulk density. Water Resour Res, 15(6), 1633-1635.

Haghverdi, A., Cornelis, W.M., Ghahraman, B. (2012) A pseudo-continuous neural network approch for developing water retention pedotransfer functions with limited data. J Hydrol, 442, 46-54. DOI: https://doi.org/10.1016/j.jhydrol

Hillel, D. (1998) Environmental Soil Physics.Academic Press, New York, USA.

IUSS Working Group W.R.B (2014) Word Reference Base for Soil Resources: International soil classification system for naming soils and creating legends for soil maps. Word soil Resour Report, (106), FAO, Rome.

Julià-Ferrer, M., Monreal, T.E., Sánchez del Corral Jiménez, A., García Meléndez, E. (2004) Constructing a saturated hydraulic conductivity map of Spain using pedotransfer functions and spatial prediction. Geoderma, 123, 257-277. DOI: https://doi.10.1016/j.geoderma.

Karandish, F., Simunek, J. (2016) A comparison of numerical and machine-learning modeling of soil water content with limited input data. J Hydrol, 543, 892-909. DOI: https://dx.doi.org/10.1016/j.hydrol.2016.11.007

Khlosi, M., Alhamdoosh, M., Douaik, A., Gabriels, D., Cornelis, W.M. (2016) Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil. Eur J Soil Sci, 67(3), 276-284. DOI: https://doi.org/10.1111/ejss.12345

Lamorski, K., Pachepsky, Y., Slawinski, C., Walczak, R.T. (2008) Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. Soil Sci Soc Am J, 72 (5), 1243-1247. DOI: https://doi.10.2136/sssaj2007.0280N

Li, X.D., Shao, M.A., Zhao, C.L. (2019) Estimating the field capacity and permanent wilting point at the regional scale for the Hexi Corridor in China using a state-space modeling approach. J Soils Sediments, 19, 3805-3816. DOI: https://doi.org/10.1007/s11368-019-02314-6

McNeill, S.J., Lilburne, L.R., Carrick, S., Webb, T.H., Cuthill, Y. (2018) Pedotransferfunctions for the soil water characteristics of New Zealand soils using S-map information. Geoderma, 326, 96-110. DOI: https://doi.org/10.1016/j.geoderma.2018.04.011

Merdum, H., Çinar, Ö., Meral, R., Apan, M. (2006) Comparison of artificial neural network and regression pedotransfer function for prediction of soil water retention and saturated conductivity. Soil Tillage Res, 90, 108-116. DOI: https://doi:10.1016/j.still.2005.08.011

Minasny, B., McBratney, A. (2002) The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci Soc Am J, 66(2), 352-361.

Morvan, X., Bruand, A., Cousin, I., Roque, J., Baran, N., Mouvet, Ch. (2004) Prédiction des propriétés de rétention en eau des sols d'un bassin versant à l'aide de fonctions de pédotransfert, influence de la densité apparente et de la teneur en élements grossiers. Étude et Gestion des Sols, 11(2), 81-99.

Nguyen, P.M., De Pue, J., Van Le, K., Cornelis, W.M. (2015) Impact of regression methods on improved effects of soil structure on soil water retention estimates. J Hydrol, 525, 598-606. DOI: https://dx.doi.org/10.1016/j.hydrol.2015.04.014

Niu, Wen-Jing., Feng, Zhong-Kai., Feng, Bao-Fei., Min, Yao-Wu., Cheng, Chun-Tian., Zhou, Jian-Zhong. (2019) Comparison of multiple lineair regression, artificial neural network, exterme learning machine and support vector machine in deriving operation rule of hydrolopower reservoir. Water, 11, 88. DOI: https://doi:10.3390/w11010088

Ostavari, Y., Sgari, K., Cornelis, W.M. (2015) Performance evaluation of pedotransfer functions to predict field capacity and permanent wilting point using UNSODA and HYPES datasets. Arid Land Resear Manag, 29, 383-398. DOI : https://doi:10.1080/15324982.1029649

Pachepsky, Y.A., Rawls, W.J., Lin, H.S. (1996) Hydropedology and pedotransfer functions.Geoderma,131,308-316.DOI : https://doi:10.1016/j.geoderma.2005.03.012

Pollaco, J.A.P. (2008) A generally applicable pedotransfert function that estimates field capacity and permanent wilting point from soil texture and bulk density. Can J Soil Sci, 88(5), 761-774.

Qiao, J.B., Zhu, Y.J., Jia, X.X., Huang, L.M., Shao, M.A. (2019) Pedotransfer functions for estimating the field capacity and permanent wilting point in the critical zone of the Loess Plateau, China. J Soils Sediments, 19,140-147. DOI: https://dx.doi.org/10.1007/s.11368-018-2036-x7

Rab, M.M., Chandra, S., Fisher, P.D., Robinson, N.J., Kitching, M., Aumann, CD., Imhof, M. (2011) Modelling and prediction of soil water contents at field capacity and permanent wilting point of dryland cropping soils. Aust J Soil Res, 49, 389-407. DOI: https://doi.10.1071/SR10160-1838-2675X/11/050389

Rawls, W., Brakensiek, D. (1982) Estimating soil water retention from soil properties. J Irrig Drain, 108(2),166-171.

Reichert, J.M., Albuquerque, J.A., kaiser, D.R., Reinert, D.J., Urach, F.L., Carlesso, R. (2009) Estimation of water retention and avaibability in soils of Rio grande do sul. R Bras Ci Solo, 33,1547-1560.

Santra, P., Das, B.S. (2008) Pedotransfert functions for soil hydraulic properties developed from a hilly watershed of Eastern India. Geoderma, 146, 439-448. DOI: https://doi:10.1016/j.geoderma.2008.06.019

Santra, P., Kumar, M., Kumawat, R.N., Painuli, D.K., Hati, K.M., Heuvelink, G.B.M., Batjes, N.H. (2018) Pedotransfert functions to estimate soil water content at field capacity and permanent wilting point in hot arid Western India. J Earth Syst Sci, 127:35. DOI : https://doi.org/10.1007/s12040-018-0937-0/j.hydrol.2012.03.036

Saxton, K.E., Rawls, W.J. (2006) Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Sci Soc Am J, 70(5), 1569–1578. DOI: https://doi.org/10.2136/sssaj2005.0117

Saxton, K.E., Rawls, W., Romberger, J.S., Papendick, R.I. (1986) Estimating generalized soil-water characteristics from texture. Soil Sci Soc Am J, 50(4),1031-1036. DOI: https://doi.org/10.2136/sssaj1986.03615995005000040039x

Sharma, S.K., Mohanty, B.P., Zhu, J. (2006) Including topography and vegetation attributes for developing pedotransfer functions. Soil Sci Soc Am J, 70(5),1430-1440. DOI: https://doi.org/10.2136/sssaj2005.0087

Skalová, J., Čistý, M., Bezák, J. (2011) Comparison of three regression models for determining water retention curves. J Hydrl Hydromech, 59(4), 275-284. DOI: https://doi:10.10.2478/v10098-011-0023-7

StatSoft (2010) STATISTICA, version 8. www.statsoft.com.

Twarakavi, N.K.C., Šïmůnek, J., Schaap, M.G. (2009) Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. Soil Sci Soc Am J,73(5), 1443-1452. DOI: https://doi:10.2136/sssaj2008.0021

Van Looy, K., Bouma, J., Herbst, M et al. (2017) Pedotransfer functions in earth system science : challenges and perspectives. Rev Geophys, 55,1199-1256. DOI: https://doi.org/10.1002/2017RG000581

Vapnik, V.N. (1995) The nature of statistical learning theory. Springer, New York.

Vapnik, V.N. (2013) The nature of statistical learning theory. Springer, New York.

Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J. (2001) Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. J Hydrol, 251, 123-150.

Zhang, Y., Schaap, M.G. (2017) Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulicparameter distributions and summary statistics, Rosetta3. J Hydrol, 547, 39-53. DOI: https://doi.org/10.1016/j.hydrol.2017.01.004