

Application of the Random Forest Model to Predict the Plasticity State of Vertisols

Yassine Al Masmoudi^{1,3}, Yassine Bouslihim^{2*}, Kaoutar Doumali¹,
Abdellah El Aissaoui³, Khalid Ibno Namr¹

¹ Laboratory of Geosciences and Environment Technics, Chouaib Doukkali University, Faculty of Science, El-Jadida, Morocco

² Laboratory of Geosciences & Environment research, Faculty of Science and Technique, Hassan I University, Settat, Morocco

³ Laboratory of Agricultural Engineering, National Institute of Agricultural Research, Settat, 26000, Morocco

* Corresponding author's e-mail: yassine.bouslihim@gmail.com

ABSTRACT

Vertisol plasticity is related to moisture content, and it requires an in-depth physicochemical characterization. This information allows us to use the land under the most adequate conditions and avoid soil physical degradation, especially its compaction. The objective of this study was to characterize the Vertisol in the Moroccan region of Doukkala-Abda and to predict soil plasticity based on the physicochemical parameters of soil, such as texture, electrical conductivity, Soil Organic Matter (SOM) and other chemical parameters for 120 samples. Determination of soil plasticity using Atterberg limits is a challenging and time-consuming method. Thus, this study aimed to develop a new model that can predict soil plasticity using the Random Forest algorithm. The soils presented homogeneity in the majority of physicochemical parameters, except a significant difference observed in the SOM and the electrical conductivity, which in turn influenced the soil plasticity state. The results showed significant and positive correlations between SOM, Soil Clay Content (SCC), Electrical Conductivity (EC), and plasticity in the Vertisol fields of the region. For the training phase, the model gave excellent results with a coefficient of determination of 0.995 and an RMSE of 0.164. Almost the same results were observed in the validation phase with a coefficient of determination of 0.974 and an RMSE of 0.361, which shows that the model succeeded in predicting plasticity in both phases. On the basis of these results, this model can be used for the plasticity prediction using other physicochemical parameters and the Random Forest Model. The prediction of soil plasticity is an important parameter to respect the timing of introducing machines/tools in the fields and avoid Vertisol degradation.

Keywords: soil plasticity, random forest, moroccan vertisols, soil degradation

INTRODUCTION

Soil, as a particular resource, plays a vital role in the sustainability and survival of civilizations. It can secure the provision of food and other essential goods (Hillel, 2009). This resource presents different physicochemical and biological properties that influence their diverse potentialities of use, such as agronomic productivity (Anderson, 2005; Resende et al., 2014). Characterization of those properties is considered as a key element for adequate soil management and conservation (Severiano, et al., 2009).

Generally, the soil is characterized by rapid degradation rates and slow regeneration processes (Van-Camp et al., 2004). This degradation can affect the productivity and the environment as well, which makes it more disastrous in case it occurs in the soil plastic state (Zuazo & Pleguezuelo, 2009). Thus, researchers have defined the plastic state as the range of soil moisture content in which soil has a plastic consistency.

The limits of this state (plastic limits) are also known as Atterberg limits in tribute to the work of Atterberg (Atterberg, 1911). In this state, soil can be sheared without forming cracks

(Campbell & O’Sullivan, 1991). The moisture content that reveals the plastic properties of the soil can be determined by the plasticity index (P_i) (Gobinath, 2016; McBride et al., 2008).

In this context, soil plasticity or plastic limits were perceived through a specific approach (Campbell & O’Sullivan, 1991), which is the water film theory. In this approach, the soil is in its plastic consistency. That is to say that the clay particles are surrounded by the water and slide each other. Moreover, if any vertical force is applied, the soil water tension can be increased and leads to plastic deformation of the soil without any regeneration. The soil water retention capacity controls the delay of this plasticity. The latter depends on the physical properties, especially the clay content (% C) and SOM (Bahmani & Palangi, 2016; Merdun et al., 2008).

In this regard, the total water storage (maximum water quantity) that can be received by the soil is determined by the measurement of several soil properties (Nyvall, 2002). SOM has a high retention capacity, and it is known as an essential factor that can be practical for use in soil water conservation, especially in the agricultural sectors (Stone & Garrison, 1940). In this respect, several studies have specified the amount of SOM that can contribute to water retention (1% SOM hold around 233,750 L/ha) (Stevenson, 1994; Mengel, 2012). In another study, Emerson (1995) indicates that SOM can contribute between 2.2% to 12.5% of the available water. This contribution is determined by soil texture, treatment, and some parameters such as organic materials and bulk density. The water content is influenced not only by the presence of SOM, but also by the clay percentage.

Thus, Vertisols, which are characterized by a significant presence of clay (>35%), can absorb more water when clay swelling occurs (Khitrov, 2016).

Vertisols, in general, are very fertile soils in many developed and developing countries (Ahmad & Mermut, 1996). These soils are often used for agricultural productions (Coulombe et al., 1996). In the American stratification (US-SEA, 1975), Vertisols are characterized by several properties (shrinking/swelling) due to the variation of the moisture content (Booltink et al., 1993). However, direct measurement of soil hydraulic properties is very important but is costly and time-consuming (Minasny & Hartemink, 2011; Rustanto et al., 2017). Thus, the

development of alternative, rapid, and inexpensive methods to estimate those properties is a good way for active and new investigations (Patil & Singh, 2016; Tomasella & Hodnett, 2004).

The majority of the environment modeling work requires quantitative soil information primarily at the regional scale (Gessler et al., 1996; Minasny et al., 2008; Hartemink & McBratney, 2008). The prediction techniques have become more popular, and they offer a faster way to estimate different soil attributes (Amanabadi et al., 2019). In this context, several studies have used the machine learning algorithms for modeling the soil classes and properties (Huang et al., 2002; Rogan et al., 2003; Sedaghat et al., 2016). These algorithms are rapid and operational in the field of soil science (Foody, 2002; Friedl & Brodley, 1997). The machine learning techniques can predict the soil properties for unvisited regions using the interrelationship with several covariates such as slope (McBratney et al., 2000), remote sensing parameters (Odeh & McBratney, 2000), soil characteristics and climatic conditions (Akpa et al., 2014). Some of the major machine language techniques used in Soil Parameters Prediction (SPP) are classification and regression trees (Breiman, 2001), k-nearest neighbor (Mansuy et al., 2014), multinomial logistic regression (Kempen et al., 2009), logistic model trees (Giasson et al., 2006), Support vector machine (Kovačević et al., 2010; Priori et al., 2014) and Random Forest model (Vågen et al., 2016). In this sense, (Rodriguez-Galiano, & Chica-Rivas, 2014) have compared some machine learning techniques, and they found that the Random Forest Method (RFM) is the most accurate and the most robust in noise and data reduction. It can operate the quantitative and qualitative datasets (Amanabadi et al., 2019). The potential of RFM in SPP has been proven by several studies (Grimm et al., 2008; Hastie et al., 2009; Vågen et al., 2016; Sreenivas et al., 2016).

The main objective of this study was to characterize the Vertisol in the Doukkala-Abda region of Morocco and to evaluate the effect of soil physical parameters on its moisture content in different fields of the region. Additionally, it aimed to examine the possibility of using the machine learning algorithms, the Random Forest method to predict the plasticity state of Vertisol in the region.

MATERIALS AND METHODS

Localization of study zone

The study zone is part of the area of the coastal Meseta in southern Morocco. It is limited in the north and northwest by the Atlantic Ocean, in the east by Oued Oum Er-rbia and the plain of Chaouia, in the south by the Mouissate collines, and in the southwest of the Abda and Rhamna massif (Fig. 1). From a geological view, the region is part of the domain of coastal Morocco (Meseta), which is framed by the chains of the Atlas and Rif.

The importance of Vertisol in the Doukkala region of Morocco

Several studies have attempted to map the geographical distribution of Vertisol in Morocco (Villar, 1953; Watteau, 1967; Bryssine, 1971). Generally, this soil type is localized in the plains of Doukkala, Chaouia, Zaer, Gharb, Loukkous, Tangier, Pre-Rif, Sais, Tadla and Haouz. In 1965, Wilbert estimated the area of Vertisol in Morocco that was about 0.2 Mha. Later on, the latter proved to be less than the real estimation of the area. Returning to previous studies, almost 6.5 Mha of soil studies that have been carried out over the last decades at different scales allowed us to estimate an area of 0.9 Mha of Vertisols distributed over several regions of Morocco. The values of the areas occupied by Vertisol across the country were cursorily estimated, as few studies have tried to

carry out an accurate estimation of Vertisols in the regional scale. As it was reported previously, the determination of the Vertisols in the Doukkala-Abda region and their plasticity was the objective of this study for managing the agricultural practicability. The different samples are represented in Figure 2. Here, the analysis shows that Vertisols represent around 29% (187 879,4 ha) in the total area of the region, which is a high percentage (Table 1).

The adequate management of this vast area can lead to improving the agricultural productivity and the economy. Moreover, it can reduce the physical degradation of the area, especially its compaction. Vertisols have high yield potential and are dedicated to intensive crop production. They are considered as the most sensitive soils to physical degradation (compaction) based on their high clay and water contents during agricultural operations.

Soil sampling and analyses

In the course of the research, 120 soil samples were collected from agricultural fields in different locations in the study area. The sampling period was between January and June 2019. These samples represent soils with different geological histories, under different climatic conditions and different soil management. Still, they almost have the same range of soil texture and OM content. Here, all the soils collected are the Vertisols of the region.



Fig. 1. Geographic localization of the Doukkala region in Morocco

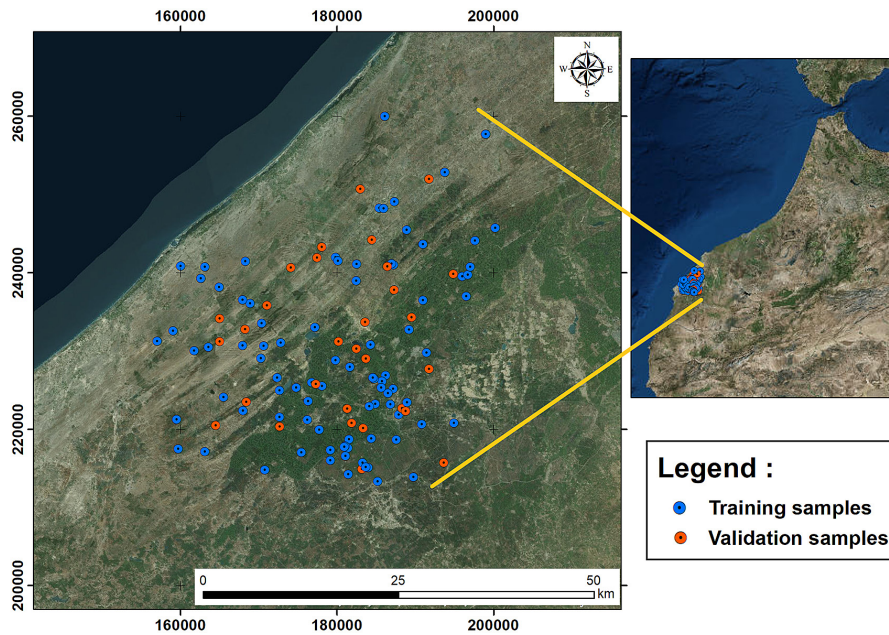


Fig. 2. Localization of the soil sampling sites

Vertisol plasticity was evaluated using the ATL method (Atterberg, 1911). A hydraulic press was used to evaluate the plasticity limits for the 120 soil samples (7 treatments and 4 replicates). Each sample (< 2 mm) was poured in a press cylinder (269.26 cm³) and humidified by tap water (150 mL). Each replicate sample was pressed up to 6 bar during different days (1st, 2nd, 3rd, 4th, 5th, 10th and 15th days) to show the variation of volume according to the moisture content. This method was developed in a previous study, in order to measure soil plasticity (Al Masmoudi et al., 2019).

The physicochemical evaluation of the Doukkala-Abda Vertisols was performed to characterize the quality of the samples. The parameters analyzed are texture (the percentage of clay, silt and sand) using Robinson Pipette, Electrical conductivity (EC) using an EM 38 conductivity meter,

Table 1. Areas of different soil types in the Doukkala region

Soil type	Soil area (ha)	Percentage area (%)
Poorly developed Calcimagnetic soils	357981	53.61
Vertisols	187879.4	28.14
Soils with iron sesquioxides	101531	15.21
Poorly developed soils	12214.5	1.83
Calcimagnetic soils	6176.8	0.93
Hydromorphic soils	1954.8	0.29

SOM using Walkey and Black method, Sodium oxide Na₂O (extracted by ammonium acetate solution), N-NO₃, N-NH₄, Boron B, Iron Fe, Manganese Mn, Zinc Zn, Copper Cu and the plasticity for the 120 samples.

Statistical analyses of data

In order to analyze the relationship between the parameters and plasticity, a statistical analysis was carried using the Statistical Package for Social Sciences (SPSS). This is done to obtain various statistical parameters such as the min, max and mean of values in each parameter of the soil samples and evaluate the difference, using the Skewness and Kurtosis statistical tests.

Random Forest Model

Random forest (RF) is considered an algorithm that allows an exploration of data, analysis and predictive modeling (Breiman, 2001). This algorithm is relatively robust to errors and outliers. That is to say, when the number of trees in the forest is large, the generalization error converges. Still, the over fitting of the training dataset cannot be a problem (Han et al., 2012). Several parameters influence the RF accuracies, such as the strength of the individual classifiers and the level of dependence between each other. Maintaining the strength of individual classifiers is ideal and optimal without increasing their correlation. The

RF model is characterized by a potential that can improve prediction using the classification and regression trees (Breiman, 2001). Indeed, the trees are constructed using the entire dataset and the splits at each node are made from the best randomly selected subset of predictors from the entire suite of input variables, which prevents over-fitting (Liaw & Wiener, 2002). The most important parameters that must be defined are the number of trees (ntree) and the number of variables randomly sampled to be tested at each node (mtry).

Development of RF model and accuracy assessment

In constructing the predictive plasticity model, the RF model was trained using 75% of data that corresponds to 90 soil samples, and the remaining samples 25%, which corresponds to 30 samples, were used for model validation. The parameters used for the development of the model do not require any preprocessing (unlike, e.g., the Linear Multiple Regression) and this is one of the advantages of this machine learning model.

The performance of the model developed by Random Forest was examined by comparing the difference between the observed and predicted plasticity coefficient, using two parameters, the coefficient of determination (R^2) and the root mean square error (RMSE).

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

$$RMSE = \sqrt{\sum_{i=1}^n [O_i - P_i]^2 / n}$$

where: O_i , P_i and \bar{O} are the observed, predicted and mean O_i value at site i , respectively; and n is the number of samples.

RESULTS

Plasticity results

The plasticity state of the different Vertisols in the region (120 samples) shows different plasticity delay (per day). The plasticity durations vary from 3 days to 8 or 12 days. On the basis of these results, the farmers can manage their Vertisol for the agricultural practicability according to the plasticity delay of their soils, considering the physicochemical parameters and the texture. Indeed, the soil plasticity state can be evaluated concerning the soil moisture content in order to timely manage the induction of machines to avoid Vertisol degradation (soil compaction). The Atterberg limits conventionally translate these differences in Vertisol behavior as a function of the water content.

Table 2. Statistical analysis of several parameters of different Vertisols of Doukkala (120 samples)

Specification	Min	Max	Mean	Ecart-type	Skewness	Kurtosis
Clay	37.00	48.80	41.65	2.62	0.67	0.05
Sand	26.80	55.20	42.16	4.82	-0.05	0.32
Silt	6.20	32.90	16.19	3.89	0.31	2.33
EC	0.04	1.40	0.80	0.44	-0.26	-1.27
SOM	0.80	3.40	1.87	0.60	0.24	-0.30
CaCO ₃	0.00	20.10	1.35	3.33	4.07	17.79
Na ₂ O	85.00	1790.00	464.63	220.75	3.36	16.28
pH	6.40	9.90	8.33	0.60	-0.15	-0.02
NT	0.06	0.17	0.10	0.02	0.90	1.37
N-NO ₃	0.09	7.59	0.94	0.91	4.03	24.12
N-NH ₄	0.01	0.70	0.24	0.14	0.70	0.47
B	0.17	1.07	0.32	0.12	3.17	16.17
Fe	2.22	33.78	12.69	6.23	1.02	1.11
Mn	1.59	34.06	6.56	4.12	2.97	16.14
Zn	0.13	3.55	0.55	0.43	3.74	20.65
Cu	0.12	1.12	0.41	0.17	1.19	3.56
Plasticity	3.00	12.00	8.08	2.18	-0.35	-0.46

Said differently, while the liquid limit characterizes the transition between the plastic state and the liquid state, the plastic limits mark the transition between the solid-state and the plastic state. Soil compaction is influenced by the moisture content. Dry soil is more difficult to compact than wet soil. If the water content increases to replace all the air occupying the pores, the soil cannot be compacted because the water is almost incompressible. The proctor compaction test is used to estimate the plasticity limits, that represent the phase when the soil is vulnerable to physical degradation (soil compaction). If the water content of a soil sample decreases, the soil successively changes from a liquid state to a plastic state, which is the most vulnerable state to compaction, for the soil is subject to irreversible deformation despite its stability, then to a solid-state.

Statistical analysis

The analytical data of the samples used for model building are presented in Table 2. There is some difference in the values of several evaluated soil parameters, as demonstrated in Table 2, the texture of different soils is almost the same, and the percentage of clay varies between 37 and 48% for the 120 soil samples. Sand and silt vary between 26.80–55.20% and 6.20–32.90%, respectively. Those soils present a percentage of SOM varied between 0.80–3.40% with a mean of 1.87%. Moreover, the electrical conductivity in the soil samples was important, reaching between 0.04–1.40 ms with a mean of 0.80 ms. The elevation of soil salinity in the coastal region is generally due to marine intrusion. Generally, data are highly skewed. Furthermore, CaCO_3 , Na_2O , N-NO_3 , B, Mn and Zn represent high kurtosis values, which means the presence of several outliers. Low kurtosis in a data set is an indicator that this data has light tails or a lack of outliers for some parameters such as clay, sand, EC, SOM, pH, and soil plasticity.

Considering the 120 samples, a high amplitude was obtained for the studied parameters. They also represent variability in different parameters. This variability of data can contribute to generating more reliable models with possible use for the soils with similar conditions, since the used values contemplate a wide range of values of the analyzed properties. Figure 3 presents the different correlations between all parameters. Here,

the clay content shows a significant correlation with the plasticity delay that goes in accordance with the high capacity of water retention. The same correlation was observed between the SOM and the plasticity is generally due to the capacity of OM to conserve water in the soil.

In the same figure, a significant correlation was found between the EC and the plasticity, which means that the soil can hold a greater quantity of water in the presence of some percentage of salinity. As it was expected, the different land uses of various areas and soil management practices, as well as the previous culture, can provide us with different values of several parameters that can, in turn, influence the plasticity timing and thus, the soil plasticity.

Random Forest results and the importance of variables

The performance of the Random Forest model was evaluated by calculating uncertainty indicators such as Coefficient of determination (R^2) and Root Mean Square Error (RMSE). For the training phase, the model gave excellent results with a coefficient of determination of 0.995 and an RMSE of 0.164 (Fig. 4). Almost the same results were observed in the validation phase with a coefficient of determination of 0.974 and an RMSE of 0.361, which shows that the model succeeded in predicting plasticity in both phases. According to these results, the model can be used for the prediction of this parameter in the case of the availability of the input data and especially the most important parameters in the prediction.

In the same sense, the main aims behind such an attempt to develop this model are to predict the plasticity delay of the Vertisols and to look for the parameters influencing the model development. Figure 5 shows the parameters that play an important role in the soil's capacity for water retention and, consequently, its duration in the plastic phase. According to the obtained results, the most relevant parameters are hydraulic conductivity, clay content and SOM. Meanwhile, the analysis of the same figure shows that the chemical parameters such as CaCO_3 , Zn, soil pH, Cu and Zn do not influence the capacity of the soil to hold the water. In other words, the model developed did not perform well due to the high variability of soil properties and low terrain variation.

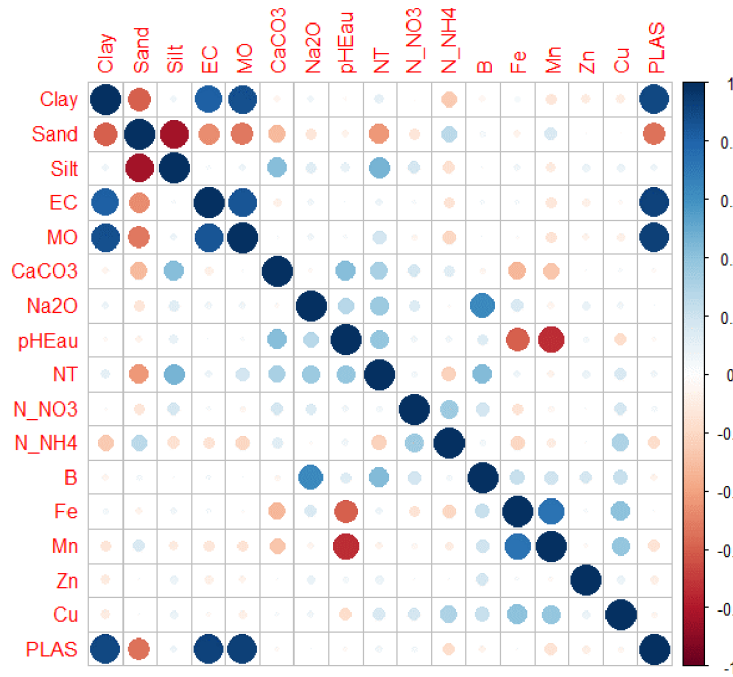


Fig. 3. Statistical correlation between physicochemical parameters of the Doukkala Vertisols

DISCUSSION

Unfortunately, the studies that predict the soil plasticity using RF are scarce to non-existent, making it impossible to compare our prediction results to other findings. Nevertheless, a comparison can be made using the research that used this type of model, especially RF, to predict other soil parameters.

In that way, Rawls et al (2003), who based their study on the USDA and NRCS soil survey characterization database, have revealed that

the soil water retention depends on the SOM and clay contents in the soil. Moreover, the increase of the SOM leads to increased water retention in the different soil textures. The latter is due to the decreases in the bulk density that affect the structure and aggregation of soil. In another system, the presence of the SOM in the O horizon can increase the soil water-holding capacity, which can, in turn, intercept the rainfall. In this regard, Hudson (1994) indicates that the increase of SOM from 0.5% to 3% leads to an increase in the water retention

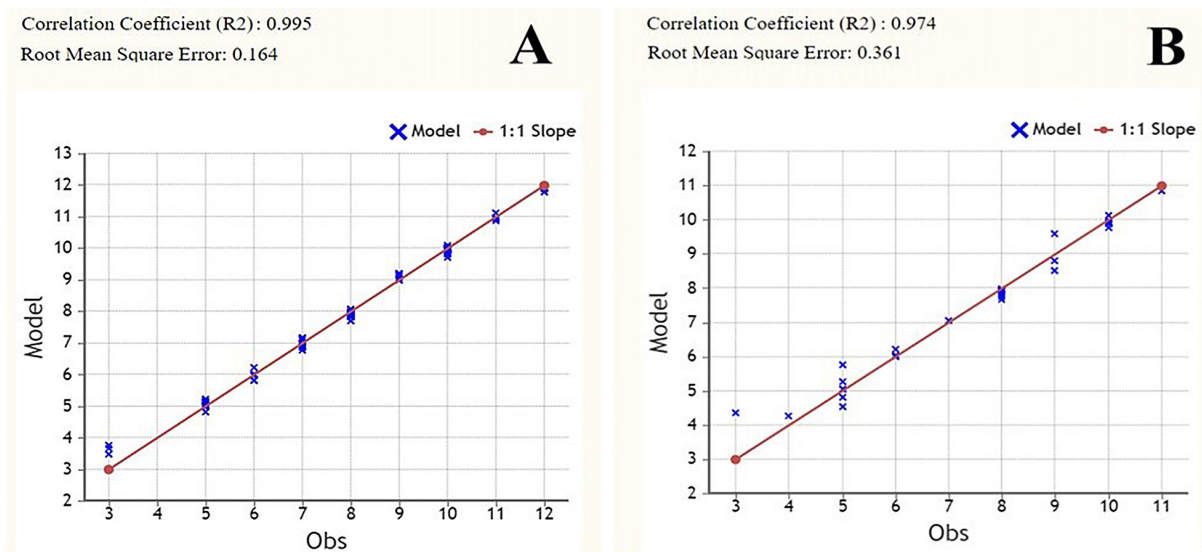


Fig. 4. Coefficient of determination (R^2) and Root Mean Square Error (RMSE) for training (A) and validation phases (B)

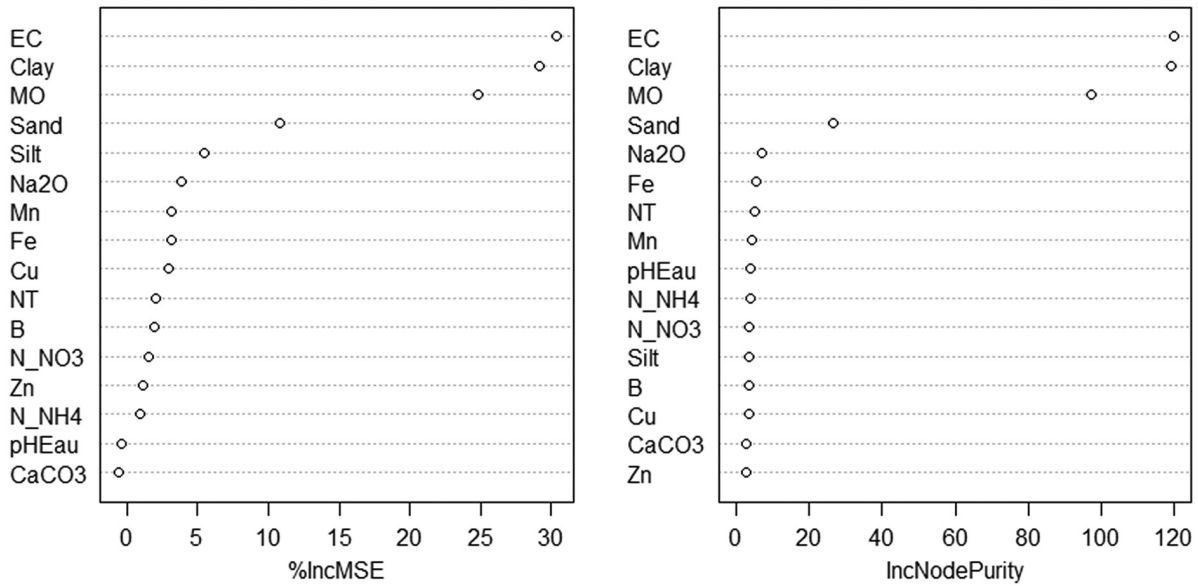


Fig. 5. Variable importance rankings of RFM model for prediction of soil plasticity (% IncMSE = percent increase in Mean Square Error)

capacity about the double in the three textures groups (sands, silt loams, and silty clay loams). On the basis of our results, there was a significant correlation between OM and plasticity delay ($R^2 = 0.949$). Here, the SOM has a strong effect on the plastic limits. This effect was especially evident while analyzing the soils of similar texture with a range of SOM contents. Still, it will be more specific when we analyze the soils from different treatments of long-term field experiments. Blanco-Canqui et al (2006) has reported that soil management significantly affects the soil consistency, and he has found significant positive correlations between SOM and plasticity period in the agricultural soils, which is in agreement with our results presented previously (Fig.3). Therefore, Keller & Dexter (2012) have developed an equation related to the plasticity limits and the clay content:

$$PL = 10.55(1.64) + 4.63 (0.241) SOM; \quad (2)$$

$$R^2 = 0.960$$

The plastic limits are affected not only by SOM but also by the soil texture, especially the clay content, as shown in the previous part (Fig.3 and Fig.5). According to the results of this study, it was a positive correlation between the clay content and plasticity delay ($R^2 = 0.9050$).

As mentioned in the Keller study (Keller et al., 2012), soil plasticity is related to the specific surface area (SSA) of the soil particles for the small particles have a higher SSA than the large

ones. In this context, we found that the plastic limits were much more strongly correlated with the clay content comparing with either the silt or sand content.

Furthermore, in some previous studies (Butler, 1955; McIntyre, 1976), authors have found that sometimes the correlation between plastic limits and the clay content may be very low. They revealed that the soil has no visible structure. The reasons behind this state may be the low SOM of these soils or the existence of micro-aggregated clay that consists of sub-plastic (Butler, 1955; Keller & Dexter, 2012; McIntyre, 1976) has come with an equation that relates the plasticity limits and the clay content:

$$PL =$$

$$= 21.28 (0.812) + 0.004 (0.0004) Clay^2; \quad (3)$$

$$R^2=0.578$$

The management of soil moisture content, especially its plastic state, can control its practicability and disregard the physical degradation of soil, particularly its compaction. The compaction can occur in the cropping cycle and affect the crop yield and performance. This physical degradation can be more severe and intense when it is applied to a Vertisol under its plastic state. In a degradation phase, the behavior of the latter changes depending on its structure and type of dominant clay. The field crop losses were due to the intensity of agricultural traffic and tillage practices.

CONCLUSIONS

The Vertisols of the Doukkala region of Morocco represent an agronomic richness with regard to their importance in agronomic productivity. Unlike the other types of soil, their physicochemical characteristics show a specific identity. The Vertisols of this region cover a significant area compared to the total area of the region. Therefore, the good management of this type of soil and its state of humidity is an essential approach for its preservation and for the sustainability of its productivity. This study evaluated the link between soil plasticity and other physicochemical parameters of Vertisols. SOM, clay content and electrical conductivity are the most relevant parameters that can contribute to soil water storage. Here, it is moisture content that can control the plasticity limits. Indeed, the realization of a model in Random Forest method can give a precise and accurate prediction of the Vertisol plasticity delay under the semi-arid conditions and identify the Vertisol parameters that influence the model development.

REFERENCES

1. Ahmad, N., & Mermut, A. (Eds.). 1996. Vertisols and technologies for their management. Elsevier.
2. Akpa, S. I., Odeh, I. O., Bishop, T. F., & Hartemink, A. E. 2014. Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal*, 78(6), 1953–1966.
3. Al Masmoudi, Y. El Aissaoui, A. and Ibno Namr, K. 2019. Evaluation of Compaction in No-Till Vertisol Field Using Methods of Cone Index and Pedotransfer Function – case of Moroccan semi-arid context . *The Journal of Agricultural Science and Technology A*, September,(9), 141–150.
4. Amanabadi, S., Vazirinia, M., Vereecken, H., Vakilian, K. A., & Mohammadi, M. H. 2019. Comparative Study of Statistical, Numerical and Machine Learning-based Pedotransfer Functions of Water Retention Curve with Particle Size Distribution Data. *Eurasian Soil Science*, 52(12), 1555–1571.
5. Anderson, Sharon. 2005. *Soils : Genesis and geomorphology*. Cambridge University Press
6. Atterberg, A. 1911. *Über die Physikalische Bodenuntersuchung Und Über die Plastizität der Tone*. *Int. Mitt. Für Bodenkunde*, 1.
7. Bahmani, O., & Palangi, S. 2016. Evaluation of pedotransfer functions for estimating the soil water retention points. *Eurasian Soil Science*, 49(6), 652–660.
8. Blanco-Canqui, H., Lal, R., Post, W. M., Izaurrealde, R. C., & Shipitalo, M. J. 2006. Organic carbon influences on soil particle density and rheological properties. *Soil Science Society of America Journal*, 70(4), 1407–1414.
9. Booltink, H. W. G., Hatano, R., & Bouma, J. 1993. Measurement and simulation of bypass flow in a structured clay soil: a physico-morphological approach. *Journal of Hydrology*, 148(1–4), 149–168.
10. Breiman, L. 2001. Random forests. *Machine learning*, 45(1), 5–32.
11. Bryssine, G. 1971. *Les Tirs du Maroc*, Les cahiers de la recherche agronomique, INRA Editions : 87–112.
12. Butler, B. E. 1955. A system for the description of soil structure and consistence in the field. *J. Aust. Inst. Agric. Sci*, 21, 239–49.
13. Campbell, D. J., & O'sullivan, M. F. 1991. The cone penetrometer in relation to trafficability, compaction, and tillage. IN: *Soil Analysis: Physical Methods*. Marcel Dekker, Inc., New York. 399–429.
14. Coulombe, C. E., Wilding, L. P., & Dixon, J. B. 1996. Overview of Vertisols: characteristics and impacts on society. In *Advances in Agronomy*, 57, 289–375.
15. Emerson, W. W. 1995. Water-retention, organic-C and soil texture. *Soil Research*, 33(2), 241–251.
16. Foody, G. M. 2002. Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1), 185–201.
17. Friedl, M. A., & Brodley, C. E. 1997. Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3), 399–409.
18. Gessler, P. E., McKenzie, N. J., & Hutchinson, M. F. 1996. Progress in soil-landscape modeling and spatial prediction of soil attributes for environmental models. In *Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Barbara, CA, 21–26.
19. Giasson, E., Clarke, R. T., Inda Junior, A. V., Merten, G. H., & Tornquist, C. G. 2006. Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. *Scientia Agricola*, 63(3), 262–268.
20. Gobinath, R., Ganapathy, G. P., Akinwumi, I. I., Kovendiran, S., Hema, S., & Thangaraj, M. 2016. Plasticity, strength, permeability and compressibility characteristics of black cotton soil stabilized with precipitated silica. *Journal of central south university*, 23(10), 2688–2694.
21. Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island–Digital soil mapping using Random Forests analysis. *Geoderma*, 146(1–2), 102–113.

22. Han, J. W., Micheline, K., & Pei, J. 2012. Data Mining: Concepts and Techniques.(Fan Ming & Meng Xiaofeng, Trans.).
23. Hartemink, A. E., & McBratney, A. 2008. A soil science renaissance. *Geoderma*, 148(2), 123–129.
24. Hastie, T., Tibshirani, R., & Friedman, J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
25. Hillel, D. 2009. The mission of soil science in a changing world. *Journal of Plant Nutrition and Soil Science*, 172(1), 5–9.
26. Huang, C., Davis, L. S., & Townshend, J. R. G. 2002. An assessment of support vector machines for land cover classification. *International Journal of remote sensing*, 23(4), 725–749.
27. Hudson, B. D. 1994. Soil organic matter and available water capacity. *Journal of soil and water conservation*, 49(2), 189–194.
28. Keller, T., & Dexter, A. R. 2012. Plastic limits of agricultural soils as functions of soil texture and organic matter content. *Soil Research*, 50(1), 7–17.
29. Kempen, B., Brus, D. J., Heuvelink, G. B., & Stoorvogel, J. J. 2009. Updating the 1: 50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, 151(3–4), 311–326.
30. Khitrov, N. B. 2016. Properties and regimes of vertisols with gilgai microtopography (a review). *Eurasian Soil Science*, 49(3), 257–271.
31. Kovačević, M., Bajat, B., & Gajić, B. 2010. Soil type classification and estimation of soil properties using support vector machines. *Geoderma*, 154(3–4), 340–347.
32. Liaw, A., & Wiener, M. 2002. Classification and regression by randomForest. *R news*, 2(3), 18–22.
33. Mansuy, N., Thiffault, E., Paré, D., Bernier, P., Guindon, L., Villemaire, P., Poirier, V & Beaudoin, A. 2014. Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. *Geoderma*, 235, 59–73.
34. McBratney, A. B., Odeh, I. O., Bishop, T. F., Dunbar, M. S., & Shatar, T. M. 2000. An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3–4), 293–327.
35. McBride, R. A., Carter, M. R., & Gregorich, E. G. 2008. Soil consistency: Upper and lower plastic limits. *Soil sampling and methods of analysis*. CRC Press, Boca Raton, FL, 761–781.
36. McIntyre, D. S. 1976. Subplasticity in Australian soils. I. Description, occurrence, and some properties. *Soil Research*, 14(3), 227–236.
37. Mengel, D. 2012. The value of soil organic matter. *K-State Agronomy Extension eUpdates*, (357).
38. Merdun, H., Meral, R., & Demirkiran, A. R. 2008. Effect of the initial soil moisture content on the spatial distribution of the water retention. *Eurasian Soil Science*, 41(10), 1098–1106.
39. Minasny, B., & Hartemink, A. E. 2011. Predicting soil properties in the tropics. *Earth-Science Reviews*, 106(1–2), 52–62.
40. Minasny, B., McBratney, A. B., & Lark, R. M. 2008. Digital soil mapping technologies for countries with sparse data infrastructures. In *Digital soil mapping with limited data*, 15–30.
41. Moussadek, R. 2012. Impacts de l’agriculture de conservation sur les propriétés et la productivité des vertisols du Maroc Central (Doctoral dissertation, Ghent University).
42. Nyvall, J. 2002. Soil water storage capacity and available soil moisture. *Water Conservation Fact-sheet no 619000–1*, Ministry of Agriculture, Food and Fisheries. British Columbia, Canada, 4.
43. Odeh, I. O., & McBratney, A. B. 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi Valley of eastern Australia. *Geoderma*, 97(3–4), 237–254.
44. Patil, N. G., & Singh, S. K. 2016. Pedotransfer functions for estimating soil hydraulic properties: A review. *Pedosphere*, 26(4), 417–430.
45. Priori, S., Bianconi, N., & Costantini, E. A. 2014. Can γ -radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine-learning methods. *Geoderma*, 226, 354–364.
46. Rawls, W. J., Pachepsky, Y. A., Ritchie, J. C., Sobecki, T. M., & Bloodworth, H. 2003. Effect of soil organic carbon on soil water retention. *Geoderma*, 116(1–2), 61–76.
47. Resende, M., Curi, N., Rezende, S. B., Corrêa, G. F., & Ker, J. C. 2014. *Pedologia: Base para distinção de ambientes*, eds. Editora UFPA.
48. Rodriguez-Galiano, V. F., & Chica-Rivas, M. 2014. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models. *International Journal of Digital Earth*, 7(6), 492–509.
49. Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., & Fischer, C. 2003. Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogrammetric Engineering & Remote Sensing*, 69(7), 793–804.
50. Rustanto, A., Booij, M. J., Wösten, H., & Hoekstra, A. Y. 2017. Application and recalibration of soil water retention pedotransfer functions in a tropical upstream catchment: case study in Bengawan Solo, Indonesia. *Journal of Hydrology and Hydromechanics*, 65(3), 307–320.

51. Schaetzl, R. J., & Anderson, S. 2005. Cambridge University Press. ISBN 0–521–81201–1.
52. Sedaghat, A., Bayat, H., & Sinigani, A. S. 2016. Estimation of soil saturated hydraulic conductivity by artificial neural networks ensemble in smectitic soils. *Eurasian Soil Science*, 49(3), 347–357.
53. Severiano, E. D. C., Oliveira, G. C. D., Curi, N., & Dias Júnior, M. D. S. 2009. Potencial de uso e qualidade estrutural de dois solos cultivados com cana-de-açúcar em Goianésia (GO). *Revista Brasileira de Ciência do Solo*, 33(1), 159–168.
54. Sreenivas, K., Dadhwal, V. K., Kumar, S., Harsha, G. S., Mitran, T., Sujatha, G., Suresh, G.J.R., Fyzee, M.A., & Ravisankar, T. 2016. Digital mapping of soil organic and inorganic carbon status in India. *Geoderma*, 269, 160–173.
55. Stevenson, F. J. 1994. Humus chemistry: genesis, composition, reactions. John Wiley & Sons.
56. Stone, J. T., & Garrison, C. S. 1940. Relationship between organic matter content and moisture constants of soils. *Soil Science*, 50(4), 253–256.
57. Tomasella, J., & Hodnett, M. 2004. Pedotransfer functions for tropical soils. *Developments in Soil Science*, 30, 415–429.
58. USSEA (United States. Science, & Education Administration). 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys (No. 436). Soil Conservation Service, US Department of Agriculture.
59. Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., & Gumbrecht, T. 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, 263, 216–225.
60. Van-Camp, L., Bujarrabal, B., Gentile, A. R., Jones, R. J., Montanarella, L., Olazabal, C., & Selvaradjou, S. K. 2004. Technical working groups established under the thematic strategy for soil protection.
61. Villar, E. H. D. 1953. Méthode de classification et analyse des sols.
62. Watteau, R. 1967. Carte de reconnaissance des sols de Zaer au 1/200 000. Inédit. Légende et tableau descriptif. Dir. Rech. Agron. Rabat.
63. Wilbert, J. 1965. Localisation géographique de la tirsification au Maroc. Les cahiers de la recherche agro-nomique, INRA Editions, 1–22.
64. Zuazo, V. H. D., & Pleguezuelo, C. R. R. 2009. Soil-erosion and runoff prevention by plant covers: a review. In *Sustainable agriculture* (pp. 785–811). Springer, Dordrecht.