Hindawi
Applied and Environmental Soil Science
Volume 2018, Article ID 9263296, 10 pages
https://doi.org/10.1155/2018/9263296

Research Article

Prediction of Soil Moisture-Holding Capacity with Support Vector Machines in Dry Subhumid Tropics

Jacob Kaingo,[1] **Siza D. Tumbo**,[1,2] **Nganga I. Kihupi**,[1] **and Boniface P. Mbilinyi**[1]

[1]*DEST, Sokoine University of Agriculture, P.O. Box 3003, Morogoro, Tanzania*
[2]*Ministry of Agriculture, P.O. Box 2182, Dodoma, Tanzania*

Correspondence should be addressed to Jacob Kaingo; jacobkaingo@gmail.com

Received 24 April 2018; Accepted 15 July 2018; Published 29 August 2018

Academic Editor: Rafael Clemente

Copyright © 2018 Jacob Kaingo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Soil moisture-holding capacity data are required in modelling agrohydrological functions of dry subhumid environments for sustainable crop yields. However, they are hardly sufficient and costly to measure. Mathematical models called pedotransfer functions (PTFs) that use soil physicochemical properties as inputs to estimate soil moisture-holding capacity are an attractive alternative but limited by specificity to pedoenvironments and regression methods. This study explored the support vector machines method in the development of PTFs (SVR-PTFs) for dry subhumid tropics. Comparison with the multiple linear regression method (MLR-PTFs) was done using a soil dataset containing 296 samples of measured moisture content and soil physicochemical properties. Developed SVR-PTFs have a tendency to underestimate moisture content with the root-mean-square error between 0.037 and 0.042 $cm^3 \cdot cm^{-3}$ and coefficients of determination (R^2) between 56.2% and 67.9%. The SVR-PTFs were marginally better than MLR-PTFs and had better accuracy than published SVR-PTFs. It is held that the adoption of the linear kernel in the calibration process of SVR-PTFs influenced their performance.

1. Introduction

Sustainability of crop yields in dry subhumid zones of marginal agricultural productivity requires integrated modelling approaches to provide the necessary feedback for adapting agrohydrological functions to changing seasonal soil moisture regimes [1]. Soil moisture-holding capacity is an important parameter for modelling moisture availability. It is a measure of the difference between moisture at field capacity and wilting point [2]. Moisture-holding capacity facilitates the description of soil hydrological processes such as drainage, infiltration, and percolation and is vital input data in models such as Soil Water Assessment Tool (SWAT) [3], and AQUACROP [4]. However, soil moisture data are generally in limited supply for tropical soils [5, 6], largely due to high costs of measurement and lack of associated equipment [6, 7].

Mathematical equations known as pedotransfer functions (PTFs), linking easily measured soil properties as input variables to soil moisture data, have been employed to bridge data gaps. With extensive development for temperate soils [8], PTFs application is fraught with specificity to calibration datasets [1] and geographic regions [8, 9]. Tropical soils have a bimodal particle-size distribution in contrast to the unimodal soils of the temperates [5, 10], with maximal weight percentage for clay- and sand-size fractions and low silt content [5]. This is suggested to impart contrasting soil hydraulic characteristics [5, 11–13], limiting transferability of PTFs for modelling processes across their statistical and pedoclimatic calibration bounds [9, 14].

Utility of PTFs necessitates validation or development of new PTFs for improved modelling outputs [9, 12]. Studies to this end for tropical soils in sub-Saharan Africa include Wosten et al. [7], Botula et al. [11], Young et al. [15], Mdemu and Mulengera [16], Mugabe [17], Obalum and Obi [18], and Mdemu [19]. All these studies have drawbacks including evaluation on small soil datasets or compiled soil databases or frequent application of the multiple linear regression method. Among the many PTF development methods, the multiple linear regression method has been highlighted to be inadequate in capturing the nonlinearity associated with moisture-holding properties [14, 20]. An insufficient data

size has been reported to be a major weakness for PTF evaluations [21–23]. Substantial uncertainty also exists with soil databases used to derive the PTFs [12], probably associated with data entry or measurement inconsistencies [5, 12].

Machine learning algorithms generally have better flexibility in mimicking the complex nonlinear pattern in the soil moisture continuum [11]. Enhanced computational efficiency of computers has spiralled the advancement of sophisticated machine learning algorithms such as artificial neural networks (ANNs) [9, 24], k-nearest neighbour [11, 21, 25], and support vector machines (SVMs) [20, 26–28]. Interest here is skewed to the SVMs because they are easier to implement than the popular ANNs [29] and have circumvented typical drawbacks associated with ANNs [20, 30, 31]. Results from ANNs are nonunique, highly dependent on the initialisation parameters, require a relatively large dataset for effective training, and often end in overfitting [30].

Support vector machines are a supervised machine learning algorithm based on statistical learning theory [32, 33], developed for data classification in [34] and later extended to solve regression problems [30, 31, 33]. The key advantage of the SVMs is structural risk minimisation over the empirical risk minimisation which checks overfitting during model development [29, 33]. Lamorski et al. [35] and Twarakavi et al. [20] pioneered the use of SVMs in the development of parametric and point PTFs, reporting improvements over ANNs. There is currently stimulated interest in the use of SVMs for PTF development [25, 26, 31], but with no evident work for sub-Saharan Africa soils. Flexibility of SVMs in incorporating limited soil data [27] would be of added benefit particularly for sub-Saharan countries, where soil data are limited but in high demand for developing sustainable farming systems. In view of this, the objective of this research was to apply support vector machines to develop pedotransfer functions for moisture-holding capacity using experimentally measured data.

2. Materials and Methods

2.1. Study Area. The study area was the Ilakala village in Kilosa District, Morogoro Region, Tanzania (Figure 1), within latitudes 7°5′30″S and 7°9′30″S and longitudes 36°50′30″E and 36°57′30″E. It has a total area of about 44 km^2. Agriculture (both cropping and livestock keeping) is the major livelihood activity in the area. The cropping system is a maize-sesame-pigeon pea small-holder system, with maize and pigeon peas as the main food crops. Sesame is a cash crop. Livestock keeping is typically undertaken by pastoralist communities of Masai and Sukuma ethnicities.

Major soil types traversing the study area are Hyperdystric Cambisol (loamic and ochric), Rhodic Acrisol (clayic, cutanic, epieutric, and profondic), Luvic Stagnic Umbrisol (endoeutric and loamic), Endogleyic Protovertic Eutric Cambisol (colluvic and ruptic), and Pellic Vertisol (ferric, humic, and mesotrophic) [36]. Many seasonal streams drain the area from the hilly regions in the southwest and western edges, feeding into River Mhenda that flows along the eastern edge of the village.

2.2. Soil Sampling and Analysis. A soil dataset of 296 samples collected between June 2014 and July 2015 was used in this study. Soil samples were taken from 100 locations at three depths (0–30 cm, 30–60 cm, and 60–100 cm). The 100 cm soil depth interval covers the complete root zone essential for available water for crop growth [37]. However, soil samples at the 60–100 cm depth interval were not taken at four sampling locations due to rockiness. Bulk soil samples were air-dried and crushed and sieved through a 2 mm sieve. Sieved soil samples were then analysed in the laboratory for particle-size distribution and organic carbon. Particle-size fractions were determined by the Bouyoucos hydrometer method [38] and separated according to the United States Department of Agriculture (USDA) particle-size classification system [39]. Organic carbon was determined by the wet oxidation method of Walkley and Black [40]. Duplicate undisturbed soil core samples in 100 cc Kopecky rings with height and diameter dimensions of 5 cm were used to determine soil moisture at field capacity (FC) and moisture content at wilting point (WP) using a pressure plate apparatus. Soil matric suctions of 30 kPa and 1500 kPa were used for FC (θ_{30}) and WP (θ_{1500}), respectively. The very soil core samples were used to determine bulk density (BD) after drying the soil core samples at 105°C for 24 hours [41].

2.3. Descriptive Statistical Analyses. The soil dataset was randomly split into a ratio of 2 : 1 for a training dataset (n = 198) and a testing dataset (n = 98), respectively. Descriptive statistics, normality tests, and correlation analyses were performed for constitutive soil variables in the dataset (BD, OC, sand, clay, and silt contents, FC, and WP) using the R statistical software [42].

2.4. PTFs Development. The training dataset was used for SVR model calibration. Input θ_{30} and θ_{1500} data were log-transformed prior to model development for both MLR and SVR methods. This was necessary for the target variables to conform to a normal distribution. Epsilon support vector regression (ε-SVR) was used for development of SVR-PTFs in the R software package e1071 [43]. Mathematical formulation of SVR is elegantly explained in the studies [20, 27, 33, 44]. Success of SVR calibration depends on three key issues: (1) selection of a suitable kernel function, (2) choice of the cost/regularisation parameter C, and (3) the "tube" insensitivity variable ε [20]. Table 1 shows some common kernel functions for SVR. The Gaussian radial basis function (RBF) kernel has been used most frequently [20, 25, 26, 35], but a linear kernel was chosen for this study because of the overfitting challenges reported with the RBF kernel [28].

The parameters C and ε are known as the hyperparameters and their optimisation determines how good the SVR model is, while "γ," "r," and "d" are kernel parameters. The parameter C determines the tolerance of the calibration prediction error and structural complexity of the SVR model. With large C values, higher penalties are assigned to the calibration error, resulting in model complexity and a computationally inefficient model with a low generalisation capability. The ε parameter controls the loss function which controls the width of the insensitive zone leading to

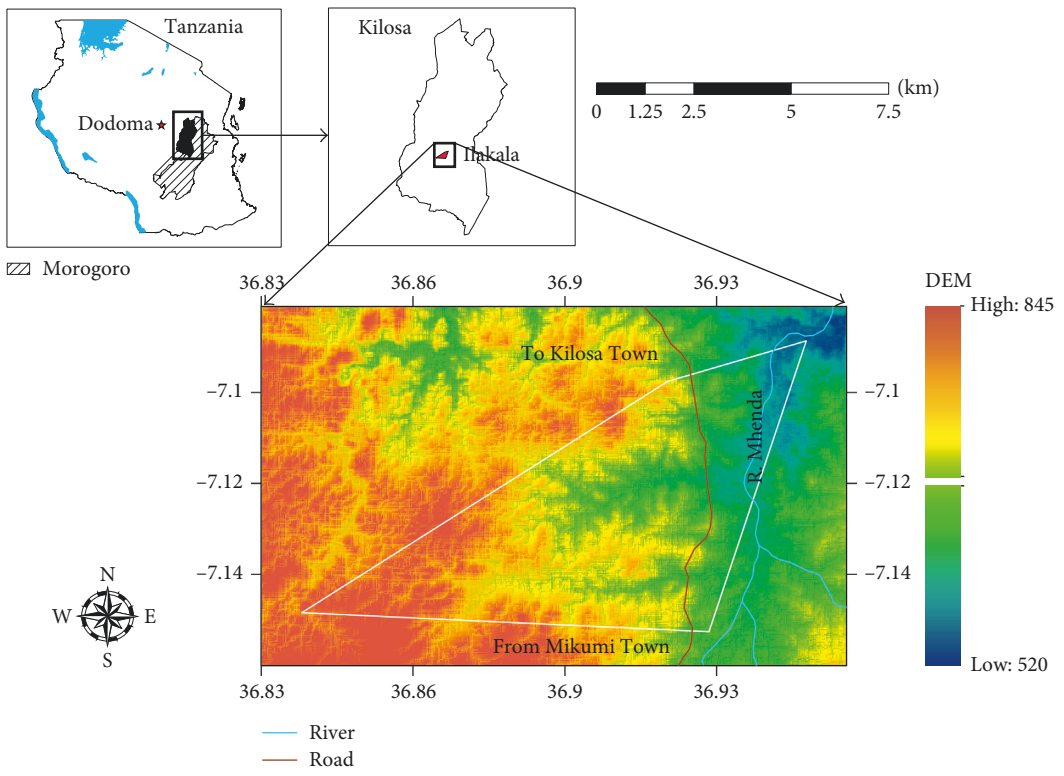


Figure 1: Location map of the Ilakala village.

Table 1: Common kernel functions and their hyperparameters in SVR.

Kernels	Functions	Parameters
Gaussian radial basis function	$e^{(-\gamma\lvert x_i^T - x_j\rvert^2)}$	C, ε, γ
Linear	$x_i^T \cdot x_j$	C, ε
Polynomial	$(\gamma x_i^T \cdot x_j + r)^d$	C, ε, r, d

minimisation of the regression risk. Large values of ε lead to smaller numbers of support vectors and poor generalisation.

The SVR calibration procedure was carried out in three steps: first, the training dataset was used to initially fit the SVR model with the linear kernel function through epsilon regression in the R software package e1071 [43]. Linear kernel functions have only two hyperparameter values that require setting, that is, the C and ε parameters. The default package C parameter value ($C = 1$) was retained for the initial fit, while the ε parameter was set to the following equation [45]:

$$\varepsilon = 3\sigma\sqrt{\left(\frac{\ln(n)}{n}\right)}, \quad (1)$$

where n is the number of records in the training dataset and σ is the standard deviation of the data.

In the second step, tuning of the SVR model hyperparameters was performed using a grid-search method with a 10-fold cross validation in 5 repeats. The grid-search method facilitates optimisation of hyperparameters by estimating the training prediction error for each set of all possible combinations of hyperparameters within the feasible feature space [20]. With insights from earlier studies [31, 35], the parameter search space was a priori set to $0.001 \le C \le 1000$ at an incremental ratio of 10 and $0 \le \varepsilon \le 0.3$ at steps of 0.001. Subsequent fine tuning was performed using a parameter search space within the neighbourhood of the best optimised hyperparameters from the second step. This process generated the best optimal hyperparameters that were ultimately used for developing the SVR-PTFs in the third step.

Multiple linear regression- (MLR-) PTFs were also developed for comparison purposes. Stepwise regression was used to develop the MLR-PTFs using the SPSS software package version 20 [46]. Both the SVR-PTFs and MLR-PTFs were then applied to the testing data to assess their validity. Performance of the developed PTFs was evaluated using the root-mean-square error (RMSE), mean error (ME), and coefficient of determination (R^2) as indicators. The RMSE, ME, and R^2 indices were calculated using (2)–(4), respectively. The RMSE and ME should ideally be close to zero, while R^2 should be close to one:

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}[y - \hat{y}], \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y - \hat{y}]^2}, \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[y - \hat{y}]^2}{\sum_{i=1}^{n}[y - \overline{y}]^2}, \quad (4)$$

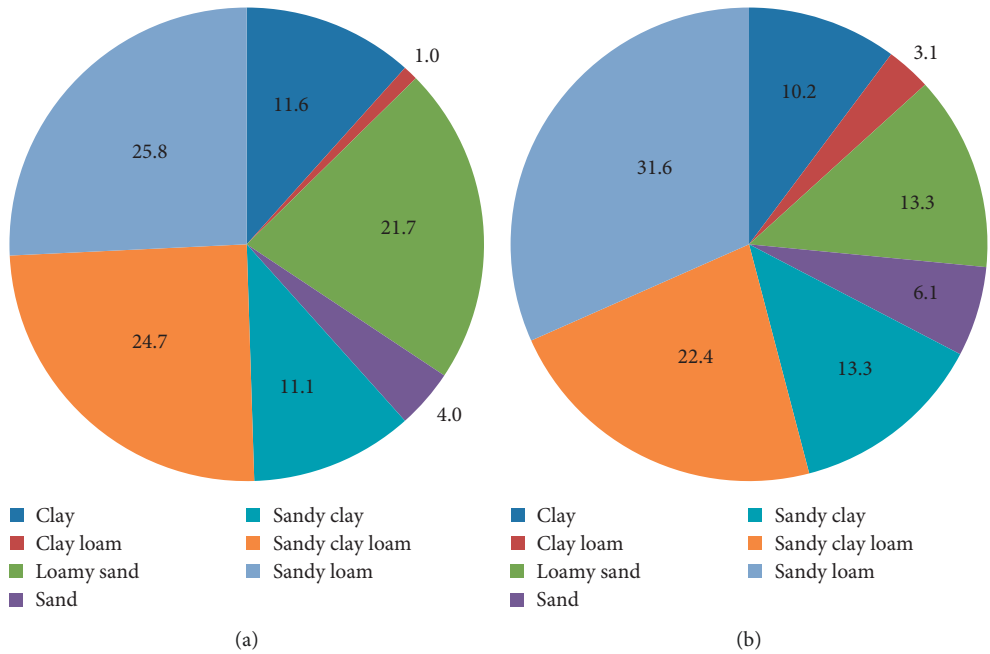


Figure 2: Percentage distribution of USDA soil texture classes between soil datasets. (a) Training dataset. (b) Testing dataset.

where y is the measured moisture content, $\hat{y}$ is the predicted moisture content, $\overline{y}$ is the mean of the measured moisture content, and n is the number of datasets.

3. Results and Discussion

3.1. Descriptive Statistics of Soil Datasets. The training and testing datasets were distributed within seven USDA soil texture classes (Figure 2). Most samples are coarse textured with more than 50% of the samples either of a sandy clay loam or a sandy loam soil textural class.

Table 2 shows the summary statistics of the training and testing datasets. Across both datasets, bulk density ranged from 1 to 1.19 g·cm^{-3}. Organic carbon ranged from 0.06% to 3.37%, and clay, sand, and silt contents ranged from 0.1% to 63.6%, 20% to 96.6%, and 1.4% to 35.4%, respectively. Moisture content at FC (θ_{30}) ranged from 0.08 to 0.48 cm^3·cm^{-3}, while moisture at WP (θ_{1500}) ranged from 0.03 to 0.39 cm^3·cm^{-3}. Mean values of training and testing datasets were similar for all soil variables. Although the skewness indices were consistent with a Gaussian distribution [47], kurtosis values were nonoptimal for an assumption of normality to be held [48].

Table 3 shows the correlation coefficients for the soil physicochemical properties on moisture content at FC (θ_{30}) and moisture content at WP (θ_{1500}). Sand and clay had a strong correlation ($r > 0.7$) but with opposite polarity for both FC and WP. Bulk density, OC, clay, and sand had highly significant correlations with moisture content at θ_{30} and θ_{1500}. Silt was poorly correlated with θ_{30} and θ_{1500} with $r < 0.07$. Organic carbon was positively correlated with moisture content at both suction extremes. Organic carbon content influences moisture retention properties due to its role in many other physical and physicochemical soil properties. Higher OC content improves soil structure and porosity, leading to increased moisture-holding capacity [3]. Organic matter also has high cation-exchange capacity and high specific surface area which enhances its moisture absorption capacity [49, 50].

Table 2: Descriptive statistics of training and testing datasets.

	Min	Max	Mean	SD	Skewness	Kurtosis
Training						
BD	1.00	1.19	1.06	0.04	1.17	1.40
OC	0.06	3.23	0.80	0.58	1.63	3.30
Clay	0.10	63.60	22.19	16.64	0.64	−0.53
Sand	20.00	96.60	64.90	16.41	−0.52	−0.35
Silt	1.40	35.40	12.90	4.89	0.87	2.75
θ_{30}	0.08	0.48	0.23	0.07	0.52	0.23
θ_{1500}	0.03	0.38	0.18	0.07	0.31	−0.13
Testing						
BD	1.00	1.16	1.05	0.03	0.97	0.88
OC	0.10	3.37	0.82	0.64	1.69	3.42
Clay	0.10	61.00	22.26	15.78	0.55	−0.64
Sand	25.60	94.60	64.67	15.97	−0.31	−0.61
Silt	2.80	24.40	13.08	5.29	0.10	−0.71
θ_{30}	0.09	0.41	0.23	0.07	0.02	−0.46
θ_{1500}	0.05	0.39	0.19	0.07	0.05	−0.26

Table 3: Pearson correlation coefficients for soil variables.

	BD (g·cc)	OC	Clay	Sand	Silt
θ_{30}	−0.46***	0.23***	0.73***	−0.76***	0.07
θ_{1500}	−0.46***	0.28***	0.77***	−0.8***	0.07

***Values are significant at a p value of 0.1%.

3.2. PTFs Development. The initial fit generated SVR models with support vectors ranging from 188 to 191 at hyperparameter settings of $C = 1$ and $\varepsilon = 0.034$ derived from (1)

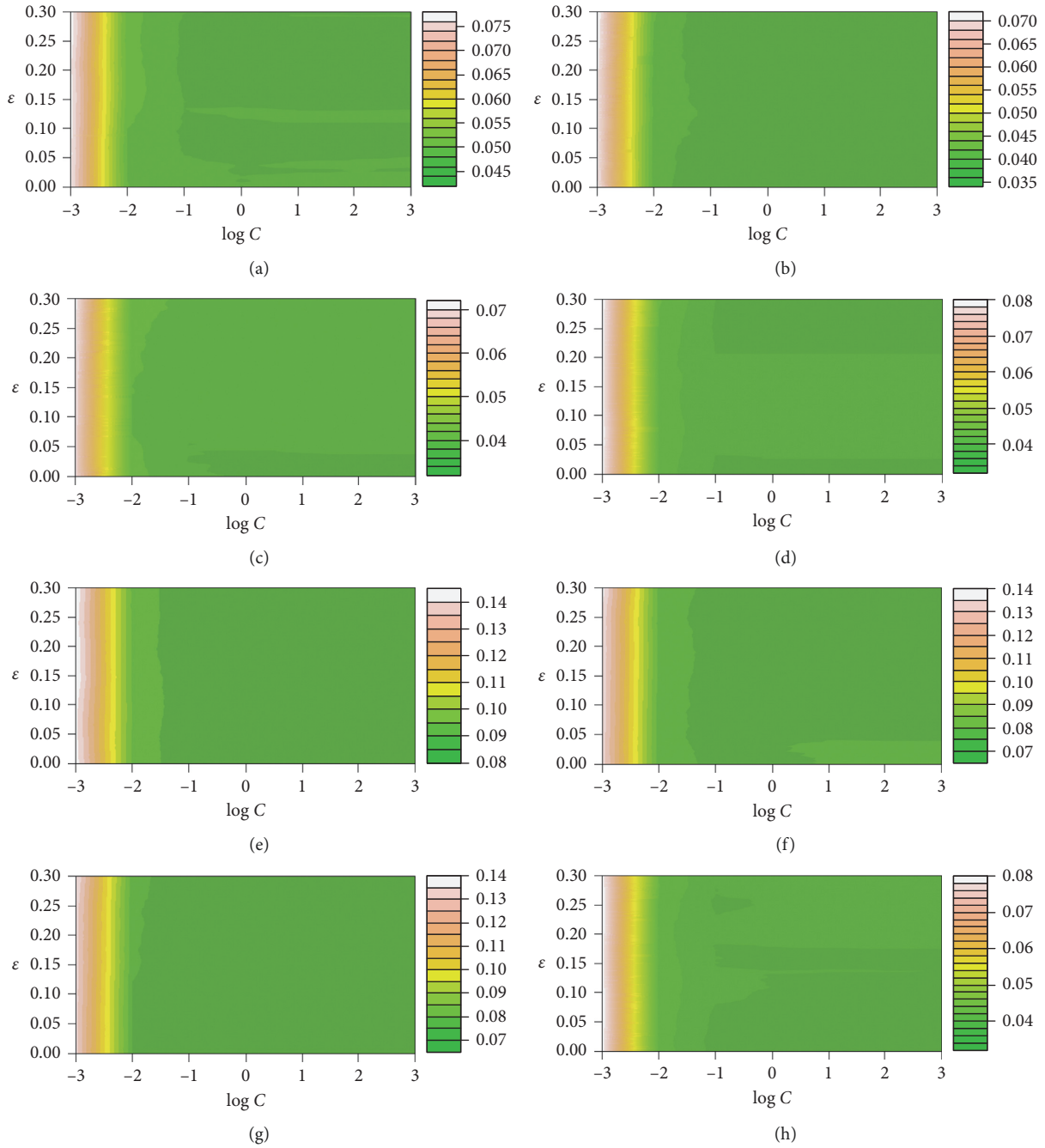


Figure 3: Sensitivity of SVR hyperparameter calibration to incremental soil predictor variables: (a) FC1, (b) FC2, (c) FC3, (d) FC4, (e) WP1, (f) WP2, (g) WP3, and (h) WP4.

(results not shown). This translated to about 95%–96% of the total support vectors used in model formulation, suggesting poor generalisation of the models with this initial choice of hyperparameters. The number of support vectors within the SVR model signifies its suitability for predictions on a new dataset. A larger proportion of support vectors lead to overfitting of the model and poor predictions on a new dataset, while a smaller proportion lead to underfitting [20]. A 50% threshold has been held as the theoretically optimal proportion of support vectors for good generalisation on new datasets [20, 28].

Figures 3(a)–3(h) show model sensitivity with varying SVR hyperparameters' combinations and increasing soil predictor variables during the coarse grid-search tuning process. The cross-validation error for FC SVR models ranged between 0.035 and 0.08 (Figures 3(a)–3(d)), respectively, corresponding to model types FC1 to FC4 (Table 4), while for WP SVR-models, the cross-validation error ranged between 0.04 and 0.14 (Figures 3(e)–3(h)) for model types WP1 to WP4. Models were most sensitive to values of C parameters, generally with lower C values ($C < 10^{-2}$) leading to higher errors for both WP and FC SVR models. The WP models were

Table 4: Calibration results of optimal hyperparameters of SVR model types.

	Predictors	C		ε		SVs		CV errors	
		1st	2nd	1st	2nd	1st	2nd	1st	2nd
FC1	Sand + clay	100	127	0.197	0.21	151	143	0.043	0.044
FC2	Clay + sand + BD	1	0.1	0	0.25	198	128	0.034	0.034
FC3	Clay + sand + BD + OC	100	96	0.006	0.022	196	189	0.034	0.034
FC4	Sand + BD	1	0.4	0.244	0.24	126	128	0.034	0.034
WP1	Sand + clay	0.1	0.35	0.199	0.217	146	138	0.080	0.081
WP2	Clay + sand + BD	1000	950	0.246	0.24	119	124	0.067	0.067
WP3	Clay + sand + BD + OC	0.1	0.65	0.15	0.15	151	149	0.065	0.065
WP4	Sand + BD + OC	10	28	0.037	0.156	183	150	0.033	0.065

most sensitive to the hyperparameters than the FC models, with the CV errors of the WP models almost twice those of FC models for the same predictor variables (Figure 3(a) versus Figure 3(e), Figure 3(b) versus Figure 3(f), and Figure 3(c) versus Figure 3(g)). However, that was not the case for the models FC4 and WP4, which showed similar CV errors because the model WP4 had an extra predictor variable (OC) than model FC4.

The pattern of higher CV errors for the WP SVR models was perhaps related to the input predictor variables in the model. Apparently, inclusion of sand as a predictor variable in WP models is counterintuitive as its adsorption of moisture at high matric suctions (i.e., 1500 kPa) is negligible. At high soil matric suctions, forces of molecular attraction (specific surface area and capillary forces) are greatly responsible for moisture retention in the soil matrix [49, 51]. Specific surface area is highly dependent on the soil particle-size distribution or texture class [52]—finer soil particle sizes have the highest specific surface area and sand the least. However, high correlation of sand with θ_{1500} (Table 3) informed its inclusion as a predictor for WP. This observed correlation could be linked to modification of surface properties of sand grains via fine-sized coatings (e. g. clay) thereby enhancing positive interactions with residual moisture.

Table 4 shows the most optimal hyperparameters from the coarse grid-search (1st) and fine grid-search (2nd) processes, with their corresponding cross-validation errors (CV errors) and number of support vectors (SVs). A lower number of SVs were evident for the SVR models after the 2nd tuning except for the FC4 model.

Table 5 shows the coefficients for the MLR-PTFs. All input variables were statistically significant. The tolerance and VIF scores indicate that the soil variables included as inputs were important predictors for moisture retention at FC and WP. A tolerance score >0.1 and VIF <10 indicate absence of multicollinearity and hence model parsimony. This implies that only the most influential predictor variables were objectively retained in the regression model. Bulk density has an inverse influence on the prediction of moisture retention. Increase in bulk density results in the destruction of the pedostructure and pore architecture, leading to a reduction in the available volume for soil moisture storage. Sand as a predictor variable had an inverse and the least influence on the moisture predictands in the MLR model. This trend could be explained by increases in soil macropores associated with sandy soils, which results in a decline in moisture retention [53]. Furthermore, sand particle fractions have a low cation-exchange capacity [54], which results in limited adsorptive sites for retaining moisture [49]. The small β coefficient for OC could have been because of the calibration of the model on a dataset with a low OC content. The average value for OC content in the training dataset was 0.8% (Table 2), which corresponds to a rating class of very low [55].

Table 5: β coefficients of MLR-PTFs for FC and WP.

Variable	β	Sig.	Tolerance	VIF*
FC				
Intercept	2.163	0	—	—
Sand	−0.013	0	0.926	1.08
Bulk density	−2.712	0	0.926	1.08
WP				
Intercept	2.916	0	—	—
Sand	−0.017	0	0.855	1.17
Bulk density	−3.493	0	0.911	1.097
OC	0.083	0.011	0.921	1.086

*VIF: variance inflation factor; Sig: significance at the 5% probability level.

Table 6: Performance indicators for SVR models with different predictors.

Model	Inputs	ME ($cm^3 \cdot cm^{-3}$)	RMSE ($cm^3 \cdot cm^{-3}$)	R^2
FC1	Sand + clay	5.53×10^{-3}	0.042	0.562
FC2	Clay + sand + BD	2.61×10^{-3}	0.038	0.643
FC3	Clay + sand + BD + OC	5.00×10^{-4}	0.037	0.663
FC4	Sand + BD	2.62×10^{-3}	0.038	0.645
WP1	Sand + clay	6.96×10^{-3}	0.045	0.546
WP2	Clay + sand + BD	3.74×10^{-3}	0.037	0.668
WP3	Clay + sand + BD + OC	2.99×10^{-3}	0.037	0.677
WP4	Sand + BD + OC	2.47×10^{-3}	0.037	0.679

3.3. Evaluation of PTFs. Table 6 shows the performance indicators for the SVR models with varying predictors. The RMSE values ranged from 0.037 $cm^3 \cdot cm^{-3}$ to 0.042 $cm^3 \cdot cm^{-3}$. These RMSE values suggest good model accuracy, given that typical RMSE values for PTFs are reported to range within 0.02 and 0.07 $cm^3 \cdot cm^{-3}$ [23]. The MEs for the developed SVR models except for the model FC3 were greater than zero, indicating a tendency to underestimate moisture at FC and WP. Coefficients of determination (R^2) were between 56.2% and 67.9% but slightly higher for the SVR models for wilting

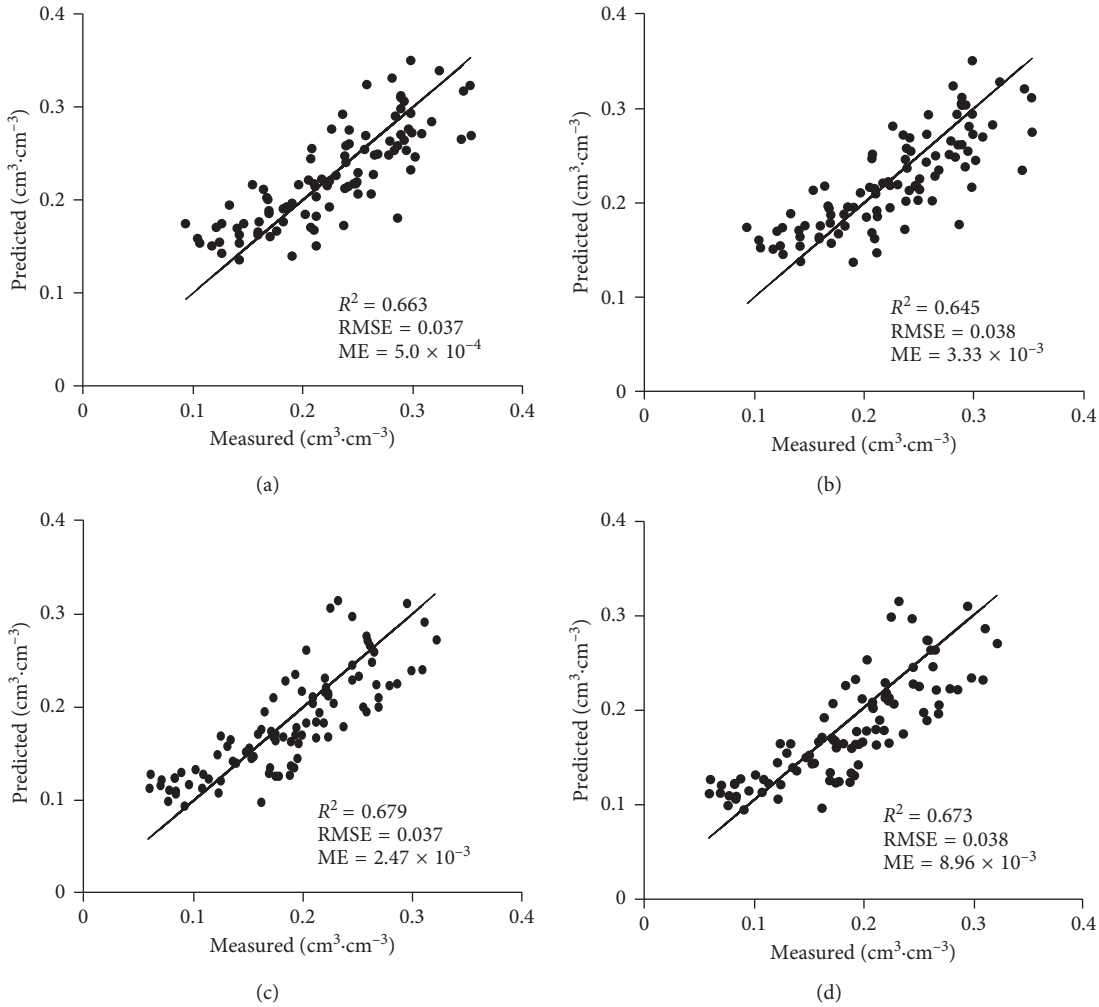


Figure 4: Unit plots of SVR- and MLR-predicted moisture content on the testing dataset. (a) SVR-FC3. (b) MLR-FC. (c) SVR-WP4. (d) MLR-WP.

point (WP2, WP3, and WP4) than the field capacity models (FC2, FC3, and FC4). The best SVR model was FC3 for moisture prediction at field capacity with sand, clay, bulk density, and organic carbon as predictors. For wilting point, the model WP4 was the best performing model with sand, bulk density, and organic carbon as predictors. The developed models explain a substantial proportion of variance of the data and provide satisfactory quantitative estimates of moisture. According to [56], models with R^2 values of 0.50 to 0.65 show good discrimination between low and high values, while those within 0.66–0.81 indicate approximate quantitative predictions, 0.82–0.90 indicate good prediction, and >0.91 indicate excellent prediction.

Unit plots of SVR- and MLR-predicted moisture content on the testing dataset are shown in Figure 4. The best performing SVR-PTFs (SVR-FC3 and SVR-WP4) were compared here. The R^2, ME, and RMSE values were marginally better for the SVR-PTFs (Figures 4(a) and 4(c)) than the developed MLR-PTFs (Figures 4(b) and 4(d)). These results are comparable to those of Nguyen et al. (2017) who also found marginal differences between MLR-PTFs and SVR-PTFs. They attributed the observed good performance of MLR-PTFs fitted by the least-squares approach to model stability which results in low variance.

Evaluation indices (R^2, ME, and RMSE) were also better at wilting point than at field capacity for both SVR and MLR models. Miháliková et al. [57] also found moisture content predictions to be more reliable at WP than at FC. The possible reason for this trend could be linked to the fact that moisture content at higher matric potentials (i.e., FC) is controlled by numerous soil factors which results in large variability within measurements. In contrast, moisture at wilting point is mainly influenced by specific surface area of the soil constituents which minimises variability in measurement values. The positive ME values indicate that the PTFs tend to underestimate moisture content. Although the ME of SVR-FC3 ($\mathrm{ME} = 5.0 \times 10^{-4}$) might suggest an unbiased model, this result was due to the deviations above and below the line of fit cancelling out.

The RMSE values for both FC and WP were $0.037\,\mathrm{cm}^3\cdot\mathrm{cm}^{-3}$ for the SVR-PTFs and $0.038\,\mathrm{cm}^3\cdot\mathrm{cm}^{-3}$ for the MLR-PTFs. The RMSE values of the SVR-PTFs developed in this study were lower than those reported in similar studies [20, 25, 26, 28, 35] (Nguyen et al., 2017), for

matric suctions at or near FC or WP. Explanation for this is not clear-cut and can only be suggested because PTF results are highly bound by datasets [14, 22, 31]. It is surmised that the observed RMSE trend is associated with the kernel function adopted in the SVR model development. This is plausible as a linear kernel was adopted in this study, while the radial basis function (RBF) kernel was used in the studies highlighted. This is corroborated by Lamorski et al. [28] who compared the linear and RBF kernels and found that the latter led to overfitting with high RMSE and poor generalisation capability on independent data when used in development of SVR-PTFs. Ben-Hur et al. [44] also observed that the use of nonlinear kernels (Gaussian RBF or polynomials) only provided marginal improvements in accuracy compared to the linear kernel. Lamorski et al. [28] attributed this to high sensitivity of SVR models to the RBF kernel width parameter (γ) (Table 1). The Gaussian RBF kernel width parameter determines the flexibility of the SVR in fitting the data, with small gamma (γ) values leading to overfitting and reduced accuracy [44]. On the contrary, it has also been shown that large γ parameters beyond an optimal threshold resulted in unrealistic R^2 values ($R^2 = 1$), which indicates overfitting [28]. Such results occur when a model starts to describe the random error in the data rather than the relationships between variables resulting in suboptimal performance outside the original training dataset [58].

Another possible explanation could be the variations in dataset characteristics used in the different studies as well as the predictors adopted for the SVR. Differences in measurement approaches and textural composition of the samples in datasets induce variability which affects the quality of PTF outputs [12, 31]. Including additional predictors to the particle-size fractions improved the accuracy of the SVR-models [20, 25, 26]. A similar trend was observed in this study. However, careful consideration is needed to avoid including difficult-to-measure soil properties as predictors.

4. Conclusions

This study was undertaken to develop SVR pedotransfer functions for estimating soil moisture-holding capacity for dry subhumid soils. Performance indices for MLR-PTFs were comparable to SVR-PTFs. The SVR-PTFs developed in this study performed slightly better than published SVR-PTFs. The linear kernel appeals for developing SVR-PTFs. However, further evaluations in this respect will be needed to establish the most optimal kernels to utilise for PTF development in view of popular application of the Gaussian radial basis function.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This research work was supported with a Ph.D. fellowship grant to Jacob Kaingo under the Soil Health Training Programme of the Alliance of a Green Revolution in Africa (AGRA) and Trans-Sec Project. The research article is based on the Ph.D. thesis of Jacob Kaingo. The authors extend gratitude to the smallholder farmers in the Ilakala village who gave permission to sample their fields. The authors appreciate the financial support from AGRA and Trans-Sec Project towards this research.

References

[1] H. Vereecken, A. Schnepf, J. W. Hopmans et al., "Modeling soil processes: review, key challenges, and new perspectives," *Vadose Zone Journal*, vol. 15, no. 5, pp. 1–57, 2016.

[2] H. Asgarzadeh, M. R. Mosaddeghi, A. A. Mahboubi, A. Nosrati, and A. R. Dexter, "Soil water availability for plants as quantified by conventional available water, least limiting water range and integral water capacity," *Plant Soil*, vol. 335, pp. 229–244, 2010.

[3] B. Tóth, M. Weynants, A. Nemes, A. Makó, G. Bilas, and G. Tóth, "New generation of hydraulic pedotransfer functions for Europe," *European Journal of Soil Science*, vol. 66, no. 1, pp. 226–238, 2015.

[4] D. Raes, P. Steduto, T. C. Hsiao, and E. Fereres, "AquaCrop—the FAO crop model to simulate yield response to water: II. Main algorithms and software description," *Agronomy Journal*, vol. 101, no. 3, pp. 438–447, 2009.

[5] B. Minasny and A. E. Hartemink, "Predicting soil properties in the tropics," *Earth Science Reviews*, vol. 106, pp. 52–62, 2011.

[6] M. G. Schaap, "Modelsfor indirect estimation of soil hydraulic properties," in *Proceeding of Encyclopedia of Hydrological Sciences*, M. Anderson, Ed., pp. 1145–1150, John Wiley & Sons, New York, NY, USA, October 2005.

[7] J. H. M. Wosten, S. J. E. Verzandvoort, J. G. B. Leenaars, T. Hoogland, and J. G. Wesseling, "Soil hydraulic information for river basin studies in semi-arid regions," *Geoderma*, vol. 195–196, pp. 79–86, 2013.

[8] P. M. Nguyen, K. Van Le, and W. Cornelis, "Using categorical soil structure information to improve soil water retention estimates of tropical delta soils," *Soil Research*, vol. 52, pp. 443–452, 2014.

[9] A. Haghverdi, W. M. Cornelis, and B. Ghahraman, "A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data," *Journal of Hydrology*, vol. 442–443, pp. 46–54, 2012.

[10] D. Condappa, S. Galle, B. Dewandel, and R. Haverkamp, "Bimodal zone of the soil textural triangle: common in tropical and subtropical regions," *Soil Science Society of America Journal*, vol. 72, no. 1, pp. 33–40, 2008.

[11] Y. D. Botula, A. Nemes, P. Mafuka, E. Van Ranst, and W. M. Cornelis, "Prediction of water retention of soils from the humid tropics by the nonparametric-nearest neighbor approach," *Vadose Zone Journal*, vol. 12, no. 2, pp. 1–17, 2013.

[12] H. Vereecken, M. Weynants, M. Javaux, Y. Pachepsky, M. G. Schaap, and M. Th. van Genuchten, "Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: a review," *Vadose Zone Journal*, vol. 9, pp. 795–820, 2010.

[13] J. H. M. Wosten, Y. A. Pachepsky, and W. J. Rawls, "Pedotransfer functions: bridging the gap between available soil data and missing soil hydraulic characteristics," *Journal of Hydrology*, vol. 251, pp. 123–150, 2001.

[14] L. F. F. Moreira, A. M. Righetto, and V. M. Medeiros, "Soilhydraulic properties estimation by using PTFs in a northeastern semiarid zone catchment, Brazil," in *Proceedings of 2nd International Environmental Modelling and Software Society Conference on Complexity and Integrated Resources Management*, Pahl-Wostl, Ed., pp. 990–995, University of Osnabruck, Osnabrück, Germany, June 2004.

[15] M. D. B. Young, J. W. Gowing, N. Hatibu, H. M. F. Mahoo, and R. W. Payton, "Assessment and development of pedotransfer functions for semi-arid sub-Saharan Africa," *Physics and Chemistry of the Earth–European Geophysical Society (B)*, vol. 24, pp. 845–849, 1999.

[16] M. V. Mdemu and M. K. Mulengera, "Using pedotransfer functions (PTFs) to estimate soil water retention characteristics (SWRCS) in the tropics for sustainable soil water management: Tanzania case study," in *Proceedings of 12th International Soil Conservation Organisation Conference on Sustainable Utilization of Global Soil and Water Resources*, J. Yuren, Ed., pp. 657–662, Beijing, China, May 2002.

[17] F. T. Mugabe, "Pedotransfer functions for predicting three points on the moisture characteristic curve of a Zimbabwean soil," *Asian Journal of Plant Science*, vol. 3, no. 6, pp. 679–682, 2004.

[18] S. E. Obalum and M. E. Obi, "Moisture characteristics and their point pedotransfer functions for coarse-textured tropical soils differing in structural degradation status," *Hydrolological Processes*, vol. 27, no. 19, pp. 2721–2735, 2013.

[19] M. V. Mdemu, "Evaluation and development of pedotransfer functions for estimating soil water holding capacity in the tropics: the case of Sokoine University of Agriculture Farm in Morogoro, Tanzania," *Journal of Geography and Geology*, vol. 7, no. 1, pp. 1–9, 2015.

[20] N. C. K. Twarakavi, J. Šimůnek, and M. G. Schaap, "Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines," *Soil Science Society of America Journal*, vol. 73, no. 5, pp. 1443–1452, 2009.

[21] A. Nemes, W. J. Rawls, and Y. A. Pachepsky, "Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties," *Soil Science Society of America Journal*, vol. 70, no. 2, pp. 327–336, 2006.

[22] A. Nemes, "Why do they keep rejecting my manuscript—do's and don'ts and new horizons in pedotransfer studies," *Agrokémia és Talajtan*, vol. 64, no. 2, pp. 361–371, 2015.

[23] Y. A. Pachepsky and W. J. Rawls, "Accuracy and reliability of pedotransfer functions as affected by grouping soils," *Soil Science Society of America Journal*, vol. 63, no. 6, pp. 1747–1757, 1999.

[24] W. A. Agyare, S. J. Park, and P. L. G. Vlek, "Artificial neural network estimation of saturated hydraulic conductivity," *Vadose Zone Journal*, vol. 6, no. 2, pp. 423–431, 2007.

[25] P. M. Nguyen, J. D. Pue, K. Van Le, and W. Cornelis, "Impact of regression methods on improved effects of soil structure on soil water retention estimates," *Journal of Hydrology*, vol. 525, pp. 598–606, 2015.

[26] M. Khlosi, M. Alhamdoosh, A. Douaik, D. Gabriels, and W. M. Cornelis, "Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil," *European Journal of Soil Science*, vol. 67, no. 3, pp. 276–284, 2016.

[27] M. Kovačević, B. Bajat, and B. Gajić, "Soil type classification and estimation of soil properties using support vector machines," *Geoderma*, vol. 154, no. 3-4, pp. 340–347, 2010.

[28] K. Lamorski, C. Slawiński, F. Moreno, G. Barna, W. Skierucha, and J. L. Arrue, "Modelling soil water retention using support vector machines with genetic algorithm optimisation," *The Scientific World Journal*, vol. 2014, Article ID 740521, 10 pages, 2014.

[29] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," 2016, http://www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf.

[30] R. M. Balabina and E. I. Lomakina, "Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data," *Analyst*, vol. 136, pp. 1703–1712, 2011.

[31] A. Haghverdi, H. S. Öztürk, and W. M. Cornelis, "Revisiting the pseudo continuous pedotransfer function concept: impact of data quality and data mining method," *Geoderma*, vol. 226-227, pp. 31–38, 2014.

[32] H. Li, W. Leng, Y. Zhou, F. Chen, Z. Xiu, and D. Yang, "Evaluation models for soil nutrient based on support vector machine and artificial neural networks," *The Scientific World Journal*, vol. 2014, Article ID 478569, 7 pages, 2014.

[33] C. H. Wu, J. M. Ho, and D. T. Lee, "Travel time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, 2004.

[34] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[35] K. Lamorski, Y. Pachepsky, C. Sławiński, and R. T. Walczak, "Using support vector machines to develop pedotransfer functions for water retention of soils in Poland," *Soil Science Society of America Journal*, vol. 72, no. 5, pp. 1243–1247, 2008.

[36] J. Kaingo and S. D. Tumbo, *Soil Mapping and Web-GIS Development for Trans-Sec Project: Final Report*, Sokoine Univesity of Agriculture, Morogoro, Tanzania, 2016.

[37] S. K. Sharma, B. P. Mohanty, and J. Zhu, "Including topography and vegetation attributes for developing pedotransfer functions," *Soil Science Society of America Journal*, vol. 70, pp. 1430–1440, 2006.

[38] G. W. Gee and J. W. Bauder, "Particle size analysis," in *Methods of Soil Analysis Part 1: Physical and Mineralogical Methods*, Monograph No. 9, A. Klute, Ed., pp. 383–411, Soil Science Society of America, Madison, WI, USA, 1986.

[39] Food and Agriculture Organisation, *Guidelines for Soil Description*, FAO, Rome, Italy, 2006.

[40] D. W. Nelson and L. E. Sommers, "Total carbon, organic carbon, and organic matter," in *Methods of Soil Analysis. Part 2—Chemical and Mineralogical Properties*, Monograph No. 9, A. L. Page, Ed., pp. 539–579, American Society of Agronomy, Madison, WI, USA, 1982.

[41] G. R. Blake and K. H. Hartge, "Bulk density," in *Methods of Soil Analysis Part 1: Physical and Mineralogical Methods*, Monograph No. 9, A. Klute, Ed., pp. 363–375, Soil Science Society of America, Madison, WI, USA, 1986.

[42] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016, http://www.R-project.org/.

[43] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, "e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien.

R package version 1.6-7," 2015, https://CRAN.R-project.org/package=e1071.

[44] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch, "Support vector machines and kernels for computational biology," *PLoS Computational Biology*, vol. 4, no. 10, article e1000173, 2008.

[45] M. Ließ, J. Schmidt, and B. Glaser, "Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches," *PLoS One*, vol. 11, no. 4, Article ID e0153673, 2016.

[46] IBM Corp., *IBM SPSS Statistics for Windows, Version 20.0.* IBM Corp., Armonk, NY, USA, 2011.

[47] D. P. Doane and L. E. Seward, "Measuring skewness: a forgotten statistic?," *Journal of Statistics Education*, vol. 19, no. 2, pp. 1–18, 2011.

[48] L. T. De Carlo, "On the meaning and use of kurtosis," *Psychological Methods*, vol. 2, no. 3, pp. 292–307, 1997.

[49] L. Khorshidi, *Soil-Water Interaction at High Soil Suction*, A thesis submitted for Award of degree of Doctor of Philosophy of Colorado School of Mines, CO, USA, 2015.

[50] T. Oztekin, B. Cemek, and L. C. Brown, "Pedotransfer functions for the hydraulic properties of layered soils," *Ziraat Fakültesi Dergisi*, vol. 24, no. 2, pp. 77–86, 2007.

[51] M. Aubertin, M. Mbonimpa, B. Bussière, and R. P. Chapuis, "A physically-based model to predict the water retention curve from basic geotechnical properties," *Canadian Geotechnical Journal*, vol. 40, no. 6, pp. 1104–1122, 2003.

[52] H. R. Fooladmand, "Estimating soil specific surface area using the summation of the number of spherical particles and geometric mean particle-size diameter," *African Journal of Agricultural Research*, vol. 6, no. 7, pp. 1758–1762, 2011.

[53] M. Tuller and D. Or, "Hydraulic conductivity of variably saturated porous media: film and corner flow in angular pore space," *Water Resources Research*, vol. 37, no. 5, pp. 1257–1276, 2001.

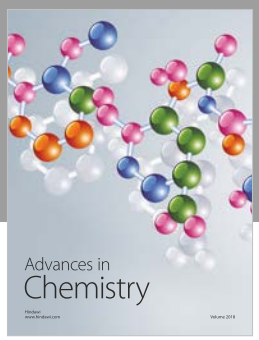
[54] International Plant Nutrition Institute (IPNI), "Cation exchange: a review," *Insights*, p. 4, 2011.

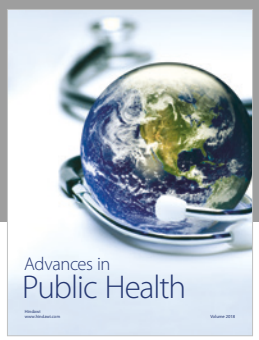
[55] D. Tsozué, P. A. Tamfuh, and S. M. N. Bonguen, "Morphology, physicochemical characteristics and land suitability in the Western Highlands of Cameroon," *International Journal of Plant & Soil Science*, vol. 7, no. 1, pp. 29–44, 2015.

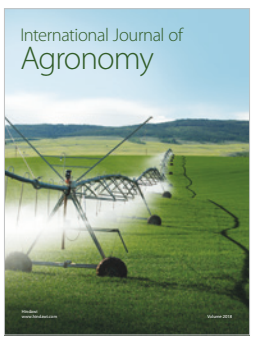
[56] A. Gholizadeh, L. Borůvka, M. M. Saberioon, J. Kozák, R. Vašát, and K. Němeček, "Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features," *Soil and Water Research*, vol. 10, no. 4, pp. 218–227, 2015.

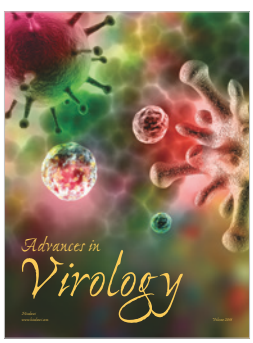
[57] M. Miháliková, M. A. Özyazıci, and O. Dengiz, "Mapping soil water retention on agricultural lands in Central and Eastern parts of the Black Sea Region in Turkey," *Journal of Irrigation and Drainage Engineering*, vol. 142, no. 12, pp. 1–9, 2016.

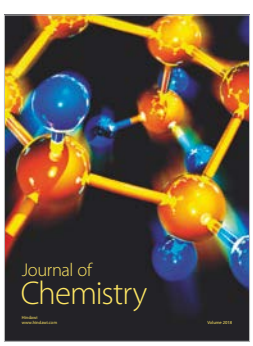
[58] M. A. Babyak, "What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models," *Psychosomatic Medicine*, vol. 66, no. 3, pp. 411–421, 2004.

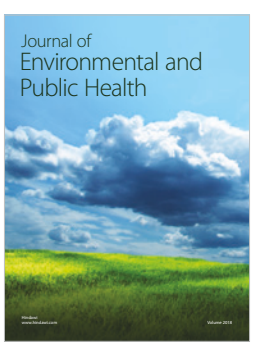


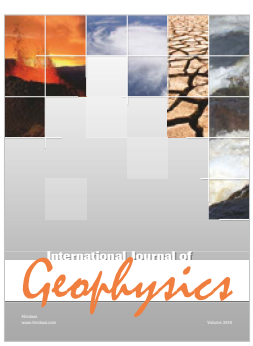


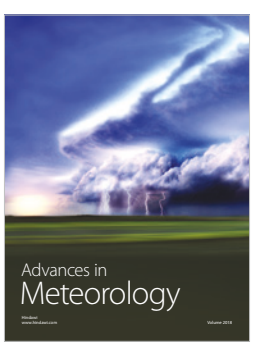


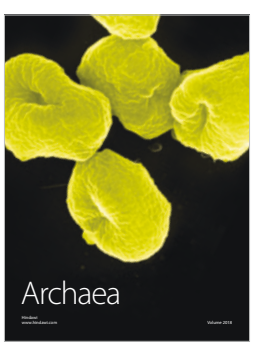


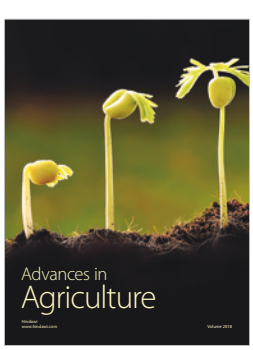












Hindawi

Submit your manuscripts at
www.hindawi.com