Scientific Research

Development of Pedotransfer Functions for Saturated Hydraulic Conductivity

John P. O. Obiero[1], Lawrence O. Gumbe[1], Christian T. Omuto[1], Mohammed A. Hassan[2], Januarius O. Agullo[1]

[1]Department of Environmental and Biosystems Engineering, University of Nairobi, Nairobi, Kenya; [2]Department of Earth and Environmental Science and Technology, Technical University of Kenya, Nairobi, Kenya.
Email: jobiero@uonbi.ac.ke

ABSTRACT

The purpose of the study was to develop pedotransfer functions for determining saturated hydraulic conductivity (K_s). Pedotransfer functions (PTFs) for predicting soil physical properties used in determining saturated hydraulic conductivity, based on moisture retention characteristics, were developed. The van Genuchten moisture retention equation was fitted to measured moisture retention properties obtained from International Soil Reference and Information Centre (ISRIC) soils data base in order to determine parameters in the equation *i.e.* saturated soil moisture content (θ_s), residual soil moisture (θ_r), air entry parameter (α) and the pore size distribution parameter (n). 457 samples drawn from the data base were used to be the maximum possible sample size that contained the measured soils characteristics data required. Using statistical regression, mathematical relationships were developed between moisture retention parameters (response variables) and appropriately selected transformed basic soil properties (predictor variables). The developed PTFs were evaluated for accuracy and reliability. It was found that pedotransfer functions developed for θ_s produced the best performance in reliability compared to the remaining parameters yielding a correlation coefficient value of coefficient of determination ($R^2 = 0.76$), $RMSE = 2.09$, $NSE = 0.75$ and $RSR = 0.5$ indicating good performance. Relatively poorest performance was obtained from the pedotransfer function developed for α which yielded a correlation coefficient, $R^2 = 0.06$, $RMSE = 0.85$ and a NSE of 0.02 reflecting the best possible equation derived for the parameter for use in predicting hydraulic conductivity. Out of the pedotransfer functions developed for each of the moisture retention parameters, the best performing PTF was identified for each parameter. The accuracy of the pedotransfer functions assessed based on R^2 were for θ_s ($R^2 = 0.80$), θ_r ($R^2 = 0.42$), α ($R^2 = 0.04$) and for n ($R^2 = 0.30$), when the variables were expressed directly in terms of the selected transformations of the basic soil properties.

Keywords: Pedotransfer Functions; Development; Evaluation

1. Introduction

Hydraulic conductivity is an important parameter that influences hydrological processes which affect flow in rivers. For instance, ground water flow is determined by saturated hydraulic conductivity (K_s). A number of formulations have been developed over the years to predict soil hydraulic conductivity from readily measurable soil properties. The need for mathematical modeling of K_s arises from the fact that insitu or laboratory measurements of hydraulic conductivity are time consuming, labour intensive and expensive as noted by [1], making it practically unlikely in reality to collect permeability data. Besides, direct measurements are also unreliable for site specific applications. The use of pe-

dotransfer functions is an alternative to determining hydraulic conductivity and involves relationships that enables K_s to be predicted using measurable soil physical and hydraulic properties. K_s serves as an input parameter in many hydrological models. There is no specific method considered most accurate in determining K_s as the method used depends on soil and environmental conditions. Furthermore, data on K_s is not readily available in most local soil survey data bases and even where they are available, the reliability is not guaranteed. There is uncertainty in the prediction of K_s using existing pedotransfer functions. Methods to develop PTFs are also unknown and not well understood especially in Africa. Well known moisture characteristic functions are based

on parameters that can be determined using moisture retention characteristics. Such data is not readily available in most soil data bases. It is essential to estimate such parameters using basic and easily available soil properties like bulk density, texture etc. Attempts to develop such equations is rare.

[2] notes that modern simulation and analysis of hydrologic processes relies heavily on the accurate and reliable description of soil water holding and transmission characteristics. Predicting K_s from basic soil properties using developed pedotransfer functions is based on analysis of the existing data sets, an approach applied in agricultural hydrology and watershed management. This research demonstrates how pedotransfer functions can be developed from a data base, with limited soils data, in predicting van Genuchten moisture retention parameters, a step in predicting K_s from readily available data on basic soil properties available from soil survey data. K_s can be indirectly predicted from soil properties using moisture retention equations like van Genuchten, Brooks Corey etc. where data on moisture retention curves is available. This technique was used in the study. Readily available data on measured K_s is rare and characterized with uncertainty on their accuracy as different methods in different environments yield varied results as observed by [3] therefore necessitating use of the indirect methods.

Three types of pedotransfer functions are recognized [4]. One category predicts soil moisture retention from basic soil properties. The other is based on point prediction of water retention characteristic as used by [5] to predict points along the moisture retention curve on Iranian soil. Another category involves prediction of parameters in models describing the θ-h-k relationship. This latter approach used in the study involved pedotransfer functions that relate the simple and easy to measure soil properties to van Genutchen moisture retention parameters, a notable gap in the science of development of pedotransfer functions. Most established pedotransfer functions for predicting soil hydraulic characteristics from continuous soil properties are based on statistical regression [6] in which the response variable, y, predicted from a number of n predictor variables, x_i yields the statistical multiple linear regression tool expressed as

$$y = a + \sum_{i=1}^{n} b_i x_i + \varepsilon, \quad i = 1, \cdots, n \quad (1)$$

where "a" is the intercept, b_i is the regression coefficient and ε is the error. The procedure to develop the pedotransfer functions involves steps which include:

The first step of statistical analysis is intended to test if each response variable distribution may be considered as normal distribution [7]. The normality of a distribution is tested by certain statistics like the Shapiro-Wilk (W-value), Skewness coefficient and Kurtosis. The shape of the distribution is determined by a measure of its Skew-

ness which is given by the equation,

$$S = \left[\frac{n}{(n-1)(n-2)} \right] \sum_{i=1}^{n} \frac{(x_i - \bar{x})^3}{s^3} \quad (2)$$

where S is Skewness coefficient, s the standard deviation, x_i the observation value, $\bar{x}$ the mean value and n the number of observations. For a normally distributed data, the skewness = 0, hence the closer the skewness coefficient is to zero (0), the better the normality of the distribution. The level of Kurtosis also provides information about the extent of the normality of the distribution and is given by,

$$K = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \right]$$
$$\times \sum_{i=1}^{n} \left[\frac{(x_i - \bar{x})^4}{\left[s^4 - (3n-1)^2 \right] / (n-2)(n-3)} \right] \quad (3)$$

The measure of Kurtosis is zero (0) for a normally distributed population. Graphical aids used to check normality includes the normal probability plot.

Mathematical relationships between potential predictor variables and the response variables are checked by detecting exponential, logarithmic, or square root tendencies. Important to eliminate is the problem of redundant information in the predictor set brought about by linear dependence between predictor variables [6]. This problem of multicollinearity can be examined by means of correlation matrix. Common statistics used in evaluating pedotransfer functions include [4]:

Multiple determination coefficient given by

$$R^2 = 1 - \frac{\sum_{1}^{N} (y_i - \hat{y}_i)^2}{\sum_{1}^{N} (y_i - \bar{y})^2} \quad (4)$$

Root mean square error given by

$$RMSE = \sqrt{\frac{\sum_{1}^{N} (y_i - \hat{y})}{N}} \quad (5)$$

Mean Error given by

$$ME = \sum_{1}^{N} \frac{(y_i - \hat{y})}{N} \quad (6)$$

Mean Absolute Error given by

$$ME = \sum_{1}^{N} \frac{|y_i - \hat{y}_i|}{N} \quad (7)$$

where y_i denotes the actual value, $\hat{y}_i$ the predicted value, and $\bar{y}$ the average of the actual value and N is the total number of observations. [8] evaluated performance of different pedotransfer functions for estimating certain soil hydraulic properties that include Available Water Capacity (AWC) using R^2, $RMSE$ and ME. [9]

used only *RMSE* in assessing pedotransfer functions developed for predicting hydraulic conductivity and water retention.

The van Genuchten water retention equation which relates moisture content (θ) to the soil matric suction (h) and is expressed as follows;

$$h(\theta) = \frac{1}{\alpha}\left[S_e^{n-1} - 1 \right]^{\frac{1}{n}} \qquad (8)$$

where

$$S_e = \left[\frac{\theta - \theta_r}{\theta_s - \theta_r} \right] \qquad (9)$$

To determine K_s, the hydraulic conductivity function used is expressed as

$$K(h) = K_s \left[\frac{\left(1 - (\alpha h)^{n-1} \left(1 + (\alpha h)^n \right)^{-m} \right)^2}{\left(1 + (\alpha h)^n \right)^{m/2}} \right] \qquad (10)$$

To determine K_s using the above equation, a relationship between hydraulic conductivity and matric suction, h is used and can be obtained through equations that relate hydraulic conductivity to permeability [10].

2. Methodology

2.1. Data Collection and Analysis

[4] noted that data from existing international soil data bases having measured soil data can be analyzed to enable prediction of hydraulic characteristics from measured soil data. Availability of measured soil hydraulic characteristics for a wide range of soils and from a large and reliable international data bases are considered prerequisite for development of pedotransfer functions. The ISRIC soils database was used to obtain measured data on soil hydraulic characteristics and basic soil properties. It was possible to obtain all the required measured data on 457 soil samples from various parts of the world. The data contained measured moisture retention curves as well as data on texture (Sand%, Clay% and Silt%). The ISRIC dataset [11] was used for this test. This dataset contains Water Retention Curve (WRC) for both topsoil (between 0 and 20 cm depth) and subsoil (between 20 and 100 cm) from many countries in the world. It has 8 levels of measured suction potential (at 0, 0.1, 0.312, 1.0, 1.99, 5.01, 2.5, and 15.8 m) (**Figure 1**). The lowest suction potential was taken as 0.01 to avoid errors with logarithmic transformations. The observed moisture retention curve obtained from measured data on moisture retention characteristics was fitted to the van Genuchten moisture retention Equation (11)
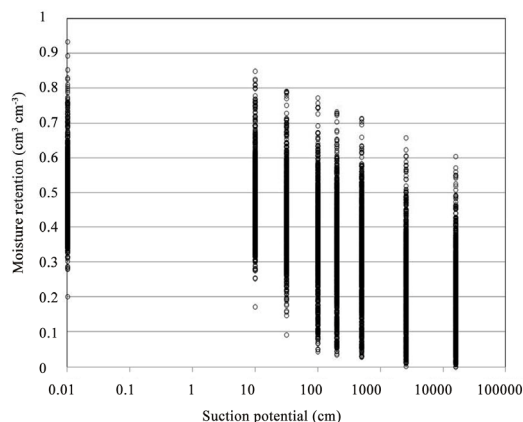
Figure 1. Measured water retention characteristics from ISRIC database.

$$\theta(h) = \theta_r + (\theta_s - \theta_r)\left[1 + (\alpha h)^n \right]^{-m} \qquad (11)$$

where $\theta(h)$ is the soil moisture content at h suction potential, θ_r is the residual moisture content, θ_s is the saturated moisture content, α is the inverse of air-entry potential, and n and m are the shape parameters for the water retention curve representing pore-size distribution index [12]. HydroMe computer programme was used to determine parameters of the van Genuchten equation (http://cran.r-project.org/web/packages/HydroMe/index.html) and is executable in *R* programme The pedotransfer functions were then developed to estimate these parameters from readily available measured basic soil properties. Relating the moisture retention parameters to the basic soil properties then makes it possible to predict the moisture retention characteristics from the basic properties and hence enabling prediction of the saturated hydraulic conductivity. It would therefore not be necessary to have measured data on moisture retention characteristics to determine K_s but instead, the basic soil properties would be used in determining the equations for moisture retention which can then be used to estimate K_s. Measurement of moisture retention characteristics is quite elaborate, tedious, time consuming and also expensive especially if a large number of sites is involved. From the sample data set of 457 samples, a sub dataset consisting of 342 (75%) samples was randomly selected for use in calibration of the pedotransfer functions to be developed while the remainder portion of 115 samples (25%) of the data was used in the validation process. In development of the pedotransfer functions, the following procedures were undertaken.

2.1.1. Evaluation of the Distribution of the Moisture Retention Parameters

A requirement in developing pedotransfer functions is that the response (dependent) variables should be normally

distributed. As a result, the moisture retention parameter of the van Genuchten equation and their possible transformations were evaluated to establish which would best approximate normal distribution. To check the extent to which the response variables would be normally distributed, statistical measures of Shapiro Wilk (W-value), Skewness Coefficient, and measures of Kurtosis were used in the assessment.

2.1.2. Statistical Regression Analysis

Statistical regression was performed between each transformed response variable and each of the basic properties and their transformations so as to establish the goodness of fit in each case based on the measure of the coefficient of determination (R^2). The purpose of this was to determine, for each response variable, the predictor variable or its transformation that gives the best quality of fit between the two when a regression is performed. This would give an indication of which predictor variable or its transformation correlates best with the response variable or it transformation.

2.1.3. Analysis of Cross Correlations

To establish the level of dependence between the selected predictor variables for each response variable best correlated to them, a cross correlation was performed between the independent variables by determining the correlation coefficient (r) between each pair of the vari ables. This kind of analysis is performed to determine if there is any correlation between two predictor variables among the set chosen for determining the multiple regression equations. It is preferable that the independent variables in a multiple linear regression equation be independent among themselves.

2.1.4. Multiple Linear Regression and Validation of Developed Equations

This was carried out between each selected transformed response variable and selected combination of predictor variables/transformation consisting of the independent variables for which the response variable best correlated as determined by the measure of coefficients of determination. The multiple coefficient of determination was determined for each equation developed to assess the accuracy of the equation. The multiple regression was carried out between response variables and the predictor variables or transformations consisting of independent variables that are not themselves significantly correlated ($r < 0.5$). After development of the pedotransfer functions using the given data set, their reliability is then tested by comparing the values of predicted parameters using the developed equations with the observed data using an independent data set. An independent data set consisting of 115 samples was used in the validation process. The

evaluation statistic used included the correlation coefficient among others.

3. Results and Discussion

3.1. Evaluation of the Distribution of the Moisture Retention Parameters

Table 1 shows the values of statistical parameters used for evaluating normality of distribution for the response variables being the van Genuchten moisture retention parameters and their proposed transformations for the data set of 457 samples. The transformations of van Genuchten parameters (response variables) that produced the best approximation to normal distribution are θ_s^{-1}, e^{θ_r}, $\sqrt{\alpha}$ and $\sqrt{n}$. This was based on measures of the Shapiro-Wilk (W-value), Skewness coefficient and Kurtosis. For illustration, **Figure 2** shows the histogram and normality plot in the normal distribution check for the saturated soil moisture content transformation (θ_s).The transformations of van Genutchen parameters (response variables) that produce the best approximation to normal distribution are θ_s^{-1}, e^{θ_r}, $\sqrt{\alpha}$ and $\sqrt{n}$. This was based on measures of the Shapiro-Wilk (W-value), Skewness coefficient and Kurtosis.

3.2. Statistical Linear Regression Analysis

The quality of fit for the linear regression between each
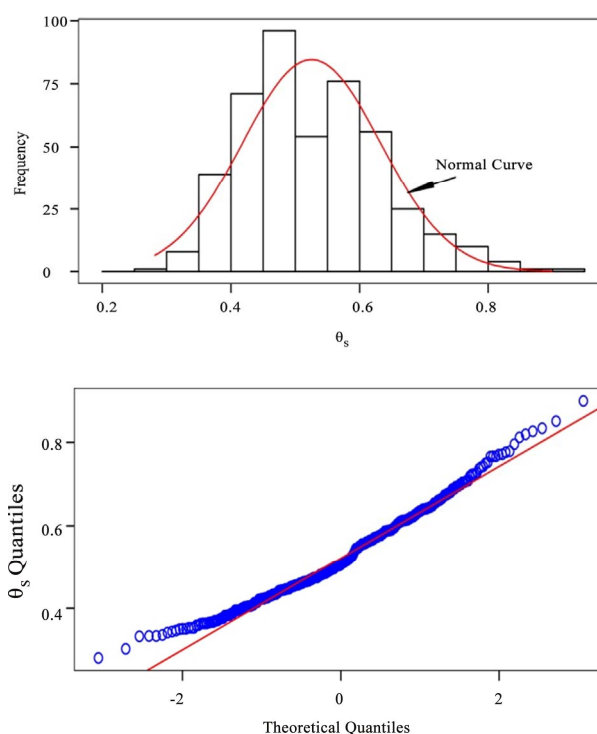
Figure 2. Check for normality of distribution of the saturated soil moisture content (θ_s) showing the histogram and normality plot.

Table 1. Normality of distribution check for parameters in the van Genuchten water retention characteristic and their transformations.

|	Kurtosis	W-value	Skewness
θ_r	0.99	0.09	−0.21
α	0.67	4.08	29.47
n	0.78	2.65	11.13
$\ln(\theta_s)$	0.99	0.04	−0.39
$\sqrt{n}$	0.99	0.28	−0.29
θ_s^{-1}	0.98	0.06	0.06
e^{θ_s}	0.96	0.82	0.66
θ_s^2	0.93	1.05	1.25
$\ln(\theta_r)$	0.75	−4.16	38.34
$\sqrt{\theta_r}$	0.93	−1.09	1.77
$1/\theta_r$	0.03	21.09	444.94
θ_r^2	0.89	1.34	2.29
e^{θ_r}	0.99	0.39	0.12
$\theta_r^{1/3}$	0.83	−2.02	5.78
$\ln(\alpha)$	0.72	−2.94	11.8
$1/\alpha$	0.03	21.17	450.93
e^{α}	0.62	5.01	44.15
α^2	0.20	13.89	238.13
$\sqrt{\alpha}$	0.95	0.91	2.27
$\text{Log}(n)$	0.98	0.49	0.94
$\sqrt{n}$	0.91	1.45	3.67
n^2	0.46	6.23	56.04
$1/n$	0.93	1.32	3.93
e^{n}	0.24	12.69	201.48

selected transformed dependent variables (moisture retention parameters) and the transformed independent variables (basic soil properties) was notably different for each predictor variable transformation. For instance, the linear regression performed between θ_s and sand yielded a quality of fit, measured by the value of R^2, to be 0.10 while that between θ_s and e^{sand} yielded $R^2 = 0.01$ (**Table 2**). The same observation was made in the case of regression performed with silt, clay, bulk density and organic carbon. Also each selected transformed variable (*i.e.* $\ln(\theta_s)$, e^{θ_r}, $\sqrt{\alpha}$, etc.) produced different values of R^2 when a linear regression was performed between each of the response variables and a transformed predictor variable for each of the variables sand, silt, clay bulk density and organic carbon. For example, a regression performed between the independent variable sand^2 and θ_s^{-1} yielded $R^2 = 0.14$ while that between sand^2 and e^{θ_r} produced $R^2 = 0.20$, hence no particular transformed variable could be said to generally give the best quality of fit with all the dependent variables whether it is sand, silt, clay, bulk density or organic carbon. Also, there is no particular response variable that could be said to yield the best quality of fit compared to others for all the independent variables or transformations to which regression was performed. The purpose of regression analysis was to identify the pair of transformed dependent and independent variables that yield the best quality of fit between them as measured by coefficient of determination (R^2) so that for each dependent variable, one is able to identify which transformations of each of the dependent variables would be used in developing the pedotransfer functions during multiple regression equation to relate each response variable and selected predictor variables and appropriate transformations. This would help establish the best possible mathematical relationship between the moisture retention parameters (response variables) and selected

Table 2. Measurement of the quality of fit based on coefficient of determination (R^2) between the moisture retention parameters (response variable) and sand with its transformation (predictor variables) in the linear regression.

	Sand	Sand^2	Sand^3	Sand^5	Sand^{10}	Sand^{15}	$\sqrt{\text{Sand}}$	ln(sand)	sand^{-1}	e$^{\text{Sand}}$
θ_s	$R^2 = 0.10$	$R^2 = 0.10$	$R^2 = 0.08$	$R^2 = 0.06$	$R^2 = 0.03$	$R^2 = 0.03$	$R^2 = 0.09$	$R^2 = 0.06$	$R^2 = 0.01$	$R^2 = 0.01$
θ_s^{-1}	$R^2 = 0.14$	$R^2 = 0.14$	$R^2 = 0.12$	$R^2 = 0.08$	$R^2 = 0.05$	$R^2 = 0.04$	$R^2 = 0.12$	$R^2 = 0.08$	$R^2 = 0.01$	$R^2 = 0.02$
$\ln(\theta_s)$	$R^2 = 0.12$	$R^2 = 0.12$	$R^2 = 0.10$	$R^2 = 0.10$	$R^2 = 0.07$	$R^2 = 0.04$	$R^2 = 0.12$	$R^2 = 0.08$	$R^2 = 0.01$	$R^2 = 0.02$
θ_R	$R^2 = 0.21$	$R^2 = 0.21$	$R^2 = 0.19$	$R^2 = 0.15$	$R^2 = 0.04$	$R^2 = 0.03$	$R^2 = 0.11$	$R^2 = 0.07$	$R^2 = 0.01$	$R^2 = 0.02$
e^{θ_r}	$R^2 = 0.20$	$R^2 = 0.20$	$R^2 = 0.18$	$R^2 = 0.13$	$R^2 = 0.07$	$R^2 = 0.05$	$R^2 = 0.18$	$R^2 = 0.13$	$R^2 = 0.02$	$R^2 = 0.03$
$\sqrt{\alpha}$	$R^2 = 0.01$	$R^2 = 0.01$	$R^2 = 0.02$	$R^2 = 0.02$	$R^2 = 0.01$	$R^2 = 0.01$	$R^2 = 0.00$	$R^2 = 0.00$	$R^2 = 0.00$	$R^2 = 0.02$
$\ln(\alpha)$	$R^2 = 0.01$	$R^2 = 0.03$	$R^2 = 0.05$	$R^2 = 0.07$	$R^2 = 0.09$	$R^2 = 0.09$	$R^2 = 0.01$	$R^2 = 0.00$	$R^2 = 0.00$	$R^2 = 0.08$
$\text{Log}(n)$	$R^2 = 0.01$	$R^2 = 0.13$	$R^2 = 0.15$	$R^2 = 0.15$	$R^2 = 0.16$	$R^2 = 0.15$	$R^2 = 0.15$	$R^2 = 0.08$	$R^2 = 0.02$	$R^2 = 0.08$
$\sqrt{n}$	$R^2 = 0.11$	$R^2 = 0.15$	$R^2 = 0.18$	$R^2 = 0.21$	$R^2 = 0.20$	$R^2 = 0.12$	$R^2 = 0.08$	$R^2 = 0.05$	$R^2 = 0.01$	$R^2 = 0.10$
n^{-1}	$R^2 = 0.07$	$R^2 = 0.08$	$R^2 = 0.09$	$R^2 = 0.08$	$R^2 = 0.09$	$R^2 = 0.06$	$R^2 = 0.06$	$R^2 = 0.04$	$R^2 = 0.01$	$R^2 = 0.03$

OJMH

basic soil properties (independent variables) in the pedotransfer function. A number of observations on the regression relationships can be made for each response variable and how it relates with each of the predictor variables and their transformations. The transformed response variable θ_s^{-1} yielded the best quality of fit (R^2 = 0.14) with the independent variables *sand* or *sand*^{2} compared to the other transformations when a regression analysis was performed. In the case of regression with silt, the best possible quality of fit was obtained with the transformation silt^{-1} (R^2 = 0.04) and so was with bulk density, BD^3 (R^2 = 0.60), $\sqrt{clay}$ or clay^2 (R^2 = 0.12), organic carbon, ln(orgC) (R^2 = 0.27). In general, the linear regression performed between θ_s^{-1} and the bulk density transformations produced best quality of fit as compared to those of silt, sand, organic carbon and clay, showing that the best relationship was obtained with bulk density followed by organic carbon, sand, clay and silt in

that order. In the case of residual moisture content, θ_r, the selected transformation variable e^{θ_r} related best with clay with a quality of fit R^2 = 0.32 obtained when regression was performed between e^{θ_r} and $\sqrt{clay}$. The poorest fit was obtained when the regression was performed with transformations of organic carbon e.g. ln(orgC) with R^2 = 0.01. α yielded the best possible quality of fit when a regression was performed with $\sqrt{clay}$ or 1/clay. $\sqrt{\alpha}$ generally showed poor quality of fit with most of the independent variables or transformations. The transformation n^{-1} also showed poor quality fit with most of the selected response variables when regression analysis was performed. The best quality fit was obtained with the transformation of sand when the linear regression was performed between n^{-1} and sand^3 (R^2 = 0.09). **Figures 3(a)-(d)** illustrates the regression between selected transformations of response variables and the predictor variables to which they best fit.
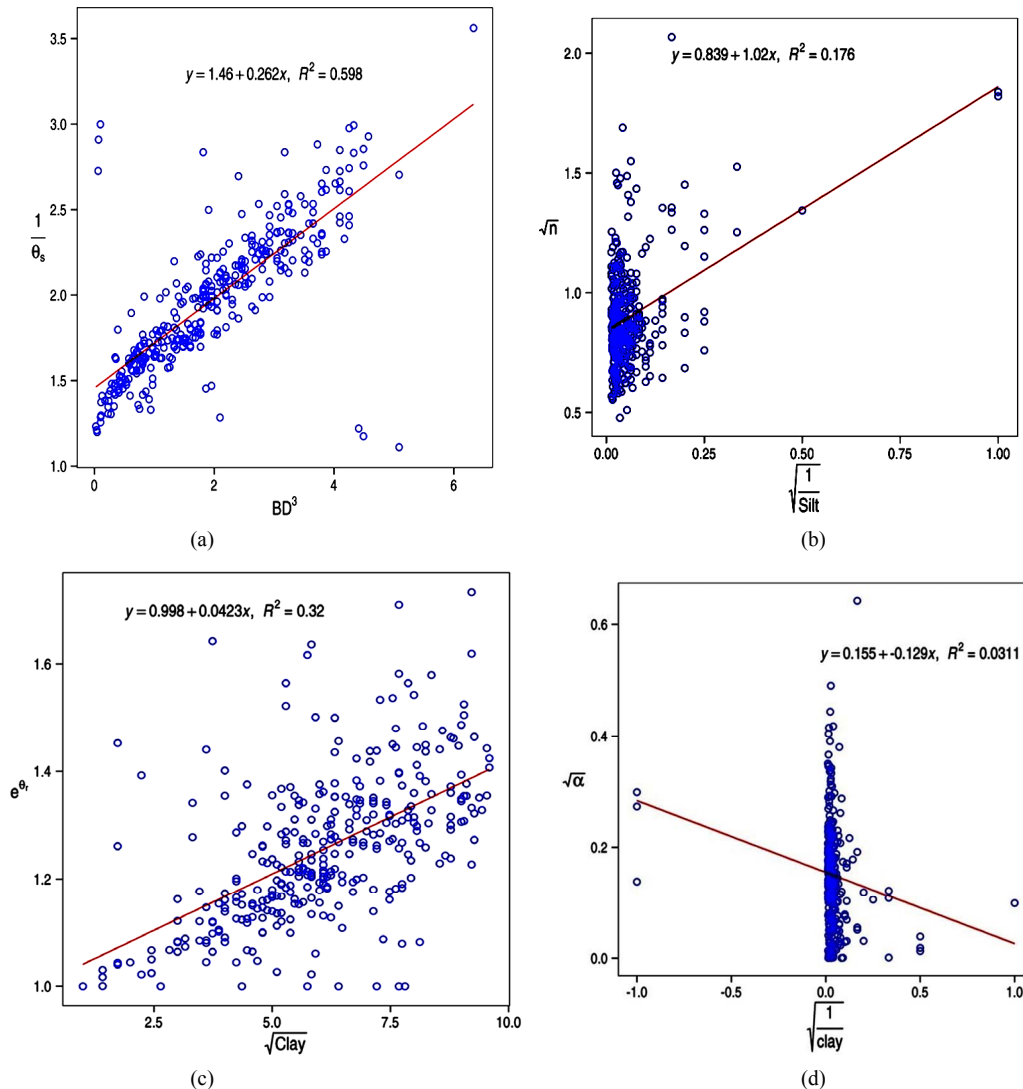
Figure 3. Response versus predictor variables correlations with corresponding best possible quality fit.

After performing linear regression between two variables involving the selected transformed response variables, associated with the moisture retention parameters, and the predictor variables associated with the basic soil properties, the following conclusions were arrived at. For the transformed response variable θ_s^{-1} the predictor variables that yielded the best possible quality of fit with it when linear regression was performed were sand, silt^{-1}, BD^3, clay^2, and ln(orgC). These variables would then be used in developing the multiple linear regression between the said response variable and the predictor variables. In the case of e^{θ_r}, the best possible fit (measured by the value of R^2) was obtained from linear regression with sand, silt^{10}, BD^2, $\sqrt{\text{clay}}$, and ln(orgC), while in the case of $\sqrt{\alpha}$ the best possible fits were obtained with e^{sand}, Silt^{-1}, BD, Clay^{-1}, and ln(orgC). In the case of $\sqrt{n}$, the best possible quality of fit was obtained with the predictor variables Sand^5, silt^{-1}, ln(clay), (orgC^2)$^{-1}$.

3.3. Multiple Regression Analysis and Accuracy of the Pedotransfer Functions Developed

Table 3 shows the quality of fit obtained when multiple linear regression was performed between each selected response variable and the transformed predictor variables to which it relates best e.g. a multiple linear regression performed between the response variable θ_s^{-1} and sand, silt^{-1}, BD^3, clay^2, ln(orgC), yielded the quality of fit measured by $R^2 = 0.63$ which is considered acceptable and therefore accurate. The resulting equation is illustrated below Equation (12). The predictor variables indicated above that related best in linear regression with the selected response variables were the ones considered in developing the pedotransfer functions using multiple linear regression.

$$\theta_s^{-1} = 1.447 + 0.003045\text{sand} - 0.1505\text{silt}^{-1} \\ + 0.2126BD^3 - 006\text{clay}^2 - 0.04542\ln(\text{orgC}) \quad (12)$$

where θ_s is the saturated soil moisture content (cm^3/cm^3), silt the % silt, sand the % sand, BD the bulk density (g/cm^3), clay the % clay, and orgC the Organic carbon content (g/kg). The qualities of fit for the remaining transformed response variables are illustrated in **Table 3** for each of e^{θ_r}, $\sqrt{\alpha}$ and $\sqrt{n}$ when the simple linear regression was done with the selected predictor variable transformations.

3.4. Analysis of Cross Correlation and Correlation Matrix

Cross correlation was done for the independent variables (transformations) used in developing multiple linear regression for each of the response variables θ_s^{-1}, e^{θ_r}, $\sqrt{\alpha}$ and $\sqrt{n}$. **Tables A1-A4** (Appendix) shows the relationships, based on correlation coefficient, between the various predictor variables used for each of the response variable transformations. In the case of θ_s^{-1} the predictor variables *sand* and $\sqrt{\text{clay}}$ appear to be highly negatively correlated with $r = -0.71$ while BD^3 and ln(orgC) are also fairly correlated in a negative sense ($r = -0.59$). For the dependent variable e^{θ_r}, the response variables BD^3 and ln(orgC) are also closely correlated in the negative sense with $r = -0.62$. Other reasonable correlations include that between e^{sand} and silt^{-2} ($r = 0.89$), BD and ln(orgC) ($r = -0.64$) in the case of predictor variables for $\sqrt{\alpha}$. The other significant correlations observed is that between sand^5 to (silt)$^{-1}$ ($r = 0.74$) and sand^5 to ln(Clay) ($r = -0.69$) in the predictor variables for $\sqrt{n}$.

3.5. Multiple Linear Regression Results

For each of the selected response variables, **Table 3** shows the various possible combinations of the predictor variables that were used in performing the multiple linear regression and also indicates the qualities of fit obtained based on the coefficient of determination R^2.

Table 3. Measurement of Quality of fit based on coefficient of determination (R^2) between the moisture retention parameters (response variables) and selected predictor variable transformations in the multiple linear regression.

	sand, silt^{-1}, BD^3, clay^2, ln(orgC)	sand, e^{BD}, sand	silt^{-1}, BD^3	$\sqrt{\text{clay}}$, BD^2, silt^{-1}	BD^3, ln(orgC)	BD^3, $\sqrt{\text{clay}}$	sand, BD^3
θ_r^{-1}	$R^2 = 0.63$	$R^2 = 0.63$	$R^2 = 0.63$	$R^2 = 0.61$	$R^2 = 0.59$	$R^2 = 0.61$	$R^2 = 0.63$
	sand, silt^{10}, BD^2, $\sqrt{\text{clay}}$, ln(orgC)	sand, $\sqrt{\text{clay}}$	$\sqrt{\text{clay}}$, silt^{10}, BD^2	clay, silt^{10}, BD^2	$\sqrt{\text{clay}}$, Silt^{10}, ln(orgC)	$\sqrt{\text{clay}}$, BD^2	$\sqrt{\text{clay}}$, ln(orgC)
e^{θ_r}	$R^2 = 0.37$	$R^2 = 0.33$	$R^2 = 0.35$	$R^2 = 0.32$	$R^2 = 0.32$	$R^2 = 0.35$	$R^2 = 0.32$
$\sqrt{\alpha}$	e^{sand},silt^{-2}, BD, clay^{-1}, ln(orgC)	e^{sand}, BD, clay^{-1}	e^{sand}, clay^{-1}, ln(orgC)	clay^{-1}, silt^{-2}, ln(orgC)	silt^{-2}, BD, clay^{-1}	BD, clay^{-1}	silt^{-1}, clay^{-1}
	0.05	0.05	0.04	0.04	0.05	0.04	0.04
	sand^5, 1/silt, ln(clay), orgC^{-2}	sand^5, orgC^{-2}	silt^{-1}, ln(clay)	ln(clay), or GC^{-2}	orgC^{-2}, BD^3	ln(clay), BD^3	sand^5, BD^3
$\sqrt{n}$	$R^2 = 0.24$	$R^2 = 0.21$	$R^2 = 0.23$	$R^2 = 0.12$	$R^2 = 0.04$	$R^2 = 0.11$	$R^2 = 0.20$

The various possible multiple linear regression equations obtained for the various independent variables are underlisted (Equations (13)-(27)). The choice of the independent variables or their transformations was based on combinations that would yield the highest possible quality of fit *i.e.* the predictor variable combinations chosen include the variable that relate very well with the response variable, hence the equations are the best possible equations that could be obtained using the variables indicated. Equations developed involving θ_s^{-1} are as follows with the measures of accuracy based on coefficient of determination (R^2) indicated:

$$\frac{1}{\theta_s} = 0.813 + 0.003534\text{sand} + 0.3099\text{e}^{BD} \quad \left(R^2 = 0.63\right) \tag{13}$$

$$\frac{1}{\theta_s} = 1.397 + 0.00317\text{sand} + 0.2451BD^3 \quad \left(R^2 = 0.63\right) \tag{14}$$

$$\frac{1}{\theta_s} = 1.397 + 0.003385\text{sand} - 0.1433\frac{1}{\text{silt}} + 0.246BD^3 \quad \left(R^2 = 0.63\right) \tag{15}$$

$$\frac{1}{\theta_s} = 1.477 + 0.003045\text{sand} - 0.1505\frac{1}{\text{silt}} + 0.212BD^3 - 9.3857E - 0.006\text{clay}^2 - 0.0452\ln(\text{orgC}) \quad \left(R^2 = 0.63\right) \tag{16}$$

The set of possible multiple linear regression equations developed involving e^{θ_r} are listed as follows:

$$\text{e}^{\theta_r} = 1.083 + 0.0382\sqrt{\text{clay}} - 5.0290E - 021\text{silt}^{10} - 00392BD^2 \quad \left(R^2 = 0.35\right) \tag{17}$$

$$\text{e}^{\theta_r} = 1.062 - 0.0007465\text{sand} + 0.03532\sqrt{\text{clay}} \quad \left(R^2 = 0.33\right) \tag{18}$$

$$\text{e}^{\theta_r} = 1.009 + 0.04099\sqrt{\text{clay}} - 4.9035E - 021\text{silt}^{10} + 0.004877\ln(\text{orgC}) \quad \left(R^2 = 0.32\right) \tag{19}$$

$$\text{e}^{\theta_r} = 1.007 + 0.03896\sqrt{\text{clay}} - 0.03928BD^2 \quad \left(R^2 = 0.35\right) \tag{20}$$

The set of possible multiple linear regression equations developed involving $\sqrt{\alpha}$ are listed as follows:

$$\sqrt{\alpha} = 0.18 - 1.0440E - 043\text{e}^{\text{sand}} - \frac{0.071381}{\text{silt}^2} - 0.02068BD - \frac{0.088661}{\text{clay}} + 0.0008447\ln(\text{orgC}) \quad \left(R^2 = 0.05\right) \tag{21}$$

$$\sqrt{\alpha} = 0.1847 - 2.0502E - 043\exp(\text{sand}) - 0.02556BD - 0.09316\frac{1}{\text{clay}} \quad \left(R^2 = 0.05\right) \tag{22}$$

$$\sqrt{\alpha} = 0.1564 - 2.0817E - 043\text{e}^{\text{sand}} - 0.08877\frac{1}{\text{clay}} + 0.004372\ln(\text{orgC}) \quad \left(R^2 = 0.04\right) \tag{23}$$

The set of possible multiple linear regression equations developed involving $\sqrt{n}$ are listed as follows:

$$\sqrt{n} = 0.9871 + 4.0051E - 011\text{sand}^5 + 0.623\frac{1}{\text{silt}} - 0.03977\ln(\text{Clay}) - \frac{0.00006391}{\text{orgC}^2} \quad \left(R^2 = 0.24\right) \tag{24}$$

$$\sqrt{n} = 1.101 + 0.8702\frac{1}{\text{silt}} - 0.07228\ln(\text{clay}) \left(R^2 = 0.24\right) \tag{25}$$

$$\sqrt{n} = 0.8617 + 8.4062E - 011\text{sand}^5 + 2.2259E - 005\left(\frac{1}{\text{orgC}^2}\right) \quad \left(R^2 = 0.21\right) \tag{26}$$

$$\sqrt{n} = 1.22 - 0.09247\ln(\text{clay}) + \frac{0.00011681}{\text{orgC}^2} \quad \left(R^2 = 0.12\right) \tag{27}$$

Based on the above indicated equations that produced the best quality of fit, equations relating the moisture retention parameters that are not transformed to the transformed basic properties were also compared to the measured (fitted) parameters and yielded the values of R^2 indicated (Equations (28) to (31)) alongside the equations.

$$\theta_s = \left(0.8103 + 0.003534\text{sand} + 0.3099\text{e}^{BD}\right)^{-1} \left(R^2 = 0.77\right) \tag{28}$$

3.6. Validation of the Developed Pedotransfer Functions to Evaluate Reliability

Table 4 shows regression equations developed for each of

$$\theta_r = \ln\left(1.083 + 0.0382\sqrt{\text{clay}} - 5.029E - 0.021\text{silt}^{10} - 0.0329BD^2\right) \quad \left(R^2 = 0.43\right) \tag{29}$$

$$n = \left(0.8617 + 8.4062E - 011\text{sand}^5 + 2.2259E - 0.005\frac{1}{\text{orgC}^2}\right)^2 \quad \left(R^2 = 0.28\right) \tag{30}$$

Table 4. Summary of equations for transformations of van Genuchten parameters and their performance in reliability.

Response variable	Equation	R^2	RMSE	RSR	NSE	ME	PBIAS
θ_s^{-1}	$\theta_s^{-1} = 0.8103 + 0.003534\text{sand} + 0.3099e^{(\text{BD})}$	0.76	2.09	0.50	0.75	0.001	0.07
θ_s^{-1}	$\theta_s^{-1} = 1.397 + 0.003176\text{sand} + 0.2451\text{BD}^3$	0.76	2.12	0.50	0.75	0.004	0.19
θ_s^{-1}	$\theta_s^{-1} = 1.477 + 0.003045\text{SAND} - 0.15051/\text{silt} + 0.2126\text{BD}^3$ $-9.3857E - 006\text{clay}^2 - 0.04542\ln(\text{orgC})$	0.74	2.11	0.51	0.74	0.009	0.44
θ_s^{-1}	$\theta_s^{-1} = 1.397 + 0.003385\text{sand} - 0.1433\text{silt}^{-1} + 0.246\text{BD}^3$	0.76	2.13	0.51	0.74	0.006	0.31
e^{θ_r}	$e^{\theta_r} = 1.083 + 0.0382\sqrt{\text{clay}} - 5.0290E - 021\text{silt}^{10} - 0.0392\text{BD}^2$	0.41	1.10	0.77	0.41	0.011	0.90
e^{θ_r}	$e^{\theta_r} = 1.062 - 0.0007465\text{sand} + 0.03532\sqrt{\text{clay}}$	0.36	1.14	0.80	0.36	0.008	0.63
e^{θ_r}	$e^{\theta_r} = 1.009 + 0.04099\sqrt{\text{clay}} - 4.9035E - 021\text{silt}^{10}$ $+0.004877\ln(\text{orgC})$	0.40	1.11	0.78	0.39	0.007	0.55
e^{θ_r}	$e^{\theta_r} = 1.007 + 0.03896\sqrt{\text{clay}} - 0.03928\text{BD}^2$	0.41	1.10	0.77	0.40	0.011	0.90
e^{θ_r}	$e^{\theta_r} = 1.003 + 0.0417\sqrt{\text{clay}} + 0.005286\ln(\text{orgC})$	0.38	1.12	0.79	0.38	0.007	0.57
$\sqrt{\alpha}$	$\sqrt{\alpha} = 0.18 - 1.0440E - 043\exp(\text{sand}) - 0.07138\text{silt}^{-1}$ $-0.02068\text{BD} - 0.088661/\text{clay} + 0.0008447\ln(\text{orgC})$	0.04	0.85	1.00	0.01	-0.012	-8.82
$\sqrt{\alpha}$	$\sqrt{\alpha} = 0.1847 - 2.0502E - 043e^{\text{sand}} - 0.02556\text{BD} - 0.09316\text{clay}^{-1}$	0.04	0.85	0.99	0.01	-0.011	-8.29
$\sqrt{\alpha}$	$\sqrt{\alpha} = 0.1564 - 2.0817E - 043e^{\text{sand}} - 0.088771/\text{clay}$ $+0.004372\ln(\text{orgC}).$	0.06	0.85	0.99	0.02	-0.013	-9.39
$\sqrt{n}$	$\sqrt{n} = 0.9871 + 4.0051E - 011\text{sand}^5 + 0.623\text{silt}^{-1}$ $-0.03977\ln(\text{Clay}) - 0.00006391/\text{orgC}^2$	0.20	1.89	0.90	0.20	-0.006	-0.62
$\sqrt{n}$	$\sqrt{n} = 1.101 + 0.87021/\text{silt} - 0.07228\ln(\text{Clay})$	0.18	1.91	0.91	0.17	-0.002	0.17
$\sqrt{n}$	$\sqrt{n} = 0.8617 + 8.4062E - 011\text{sand}^5 + 2.2259E - 0051/\text{orgC}^2$	0.20	1.88	0.89	0.20	-0.005	-0.59
$\sqrt{n}$	$\sqrt{n} = 1.22 - 0.09247\ln(\text{Clay}) + 0.00011681/\text{orgC}^2$	0.08	2.02	0.96	0.08	-0.002	0.19

$$\alpha = \left(0.1564 - 2.0817E - 043e^{\text{sand}} - 0.08877\frac{1}{\text{clay}} + 0.004372\text{Ln}(\text{orgC}) \right)^2 \quad \left(R^2 = 0.04 \right) \tag{31}$$

the selected transformations of the moisture retention parameters of the van Genutchen moisture retention equation and measures of their performance in validation based on the statistical measures indicated. The transformed model parameter that yielded the best reliability is the saturated soil moisture content θ_s^{-1} (considering the measure of coefficient of determination ($R^2 = 0.76$). This indicates that θ_s is reasonably well predicted by the pedotransfer function developed. Prediction of transformed variable $\sqrt{\alpha}$ yielded the lowest possible value for the selected measure for the quality of fit ($r = 0.25$) reflecting the lowest level of reliability compared to the other pedotransfer functions developed for moisture retention parameters under consideration. The transformed variable for residual moisture content e^{θ_r} yielded modest performance with the best possible value of $R^2 = 0.41$.

[4] notes that continuous pedotransfer functions can be applied in the case of more site specific applications, where measured data is available, since they do not provide site specific information. The pedotransfer functions predict soil hydraulic characteristics for broadly defined textural classes. **Figures 4-7** illustrates the performance of the developed pedotransfer functions for the transformed response variables based on analysis by linear regression.

4. Conclusion and Recommendation

The performance of developed pedotransfer functions for estimating moisture retention parameters in the van Genuchten moisture retention equation varied for the individual moisture retention parameters. Equations deve-

$$\frac{1}{\theta_s} = 0.813 + 0.0035\,\text{sand} + 0.3099\,e^{BD}$$
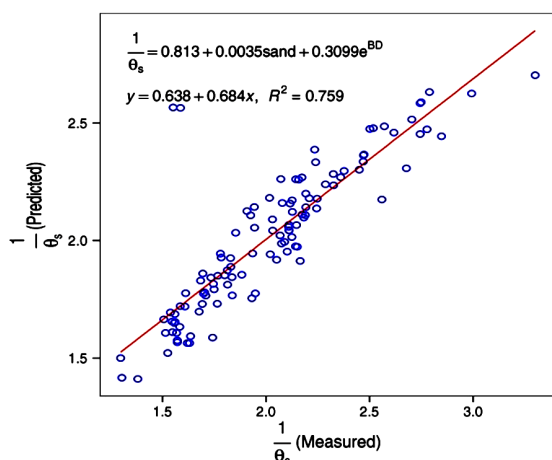$$y = 0.638 + 0.684x, \quad R^2 = 0.759$$

Figure 4. Predicted value of transformed variable θ_s^{-1} versus measured by correlation for a developed equation that yielded the best possible quality of fit.
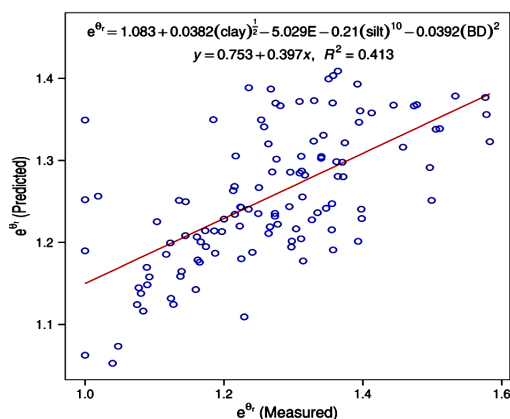

$$e^{\theta_r} = 1.083 + 0.0382(\text{clay})^{\frac{1}{2}} - 5.029E - 0.21(\text{silt})^{10} - 0.0392(BD)^2$$
$$y = 0.753 + 0.397x, \quad R^2 = 0.413$$

Figure 5. Predicted value of transformed variable e^{θ_r} versus measured by correlation for a developed equation that yielded the best possible quality of fit.

$$\sqrt{\alpha} = 0.156 - 2.0817E - 43e^{\text{sand}} - 0.0888(\text{clay})^{-1} + 0.00437\,LN(\text{orgC})$$
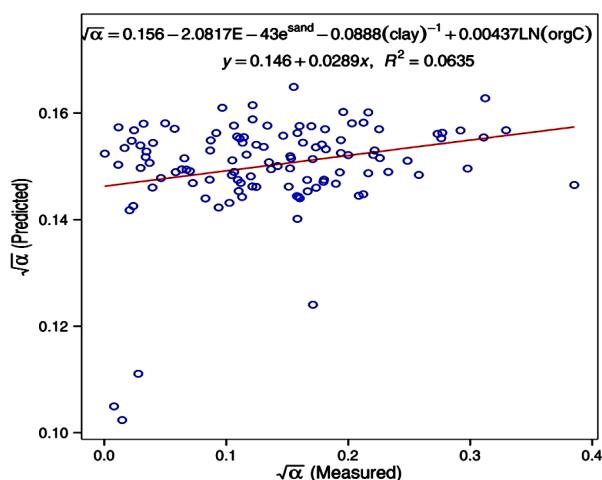$$y = 0.146 + 0.0289x, \quad R^2 = 0.0635$$

Figure 6. Predicted value of transformed variable $\sqrt{\alpha}$ versus measured by correlation for a developed equation that yielded the best possible quality of fit.

$$\sqrt{n} = 0.8617 + 8.4062E - 11(\text{sand})^5 + 2.2259E - 5(\text{orgC0})^{-2}$$
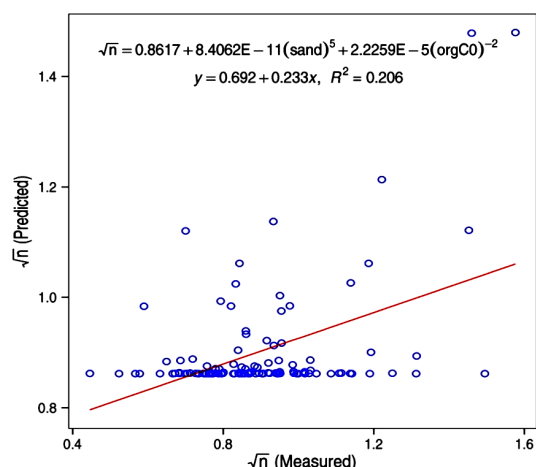$$y = 0.692 + 0.233x, \quad R^2 = 0.206$$

Figure 7. Predicted value of transformed variable $\sqrt{n}$ versus measured by correlation for a developed equation that yielded the best possible quality of fit.

loped for predicting saturated soil moisture content (θ_s) yielded the best performance in model validation with the highest possible value of coefficient of determination $R^2 = 0.76$, *RMSE* = 2.09 and a *NSE* of 0.75. The relatively poorest performance was observed with the parameter "α" with the best possible value of R^2 being 0.06 indicating unsatisfactory performance and a *RMSE* value of 0.85. The remaining parameters (θ_r and "n") registered performance intermediate between the aforementioned two extremes. It is therefore evidently not possible to obtain excellent performance in pedotransfer functions for all the parameters. The equations derived from these parameters would be the best possible obtainable for estimating hydraulic conductivity. Continued effort should be made in developing more pedotransfer functions to provide users with more alternative equations and especially those PTFs that use the very basic soil properties available in many local soils data bases.

5. Acknowledgements

REFERENCES

[1] H. R. Fooladmad, "Pedotransfer Functions for Point Estimation of Soil Moisture Characteristic Curve in Some Iranian Soils," *African Journal of Agricultural Research*,

Vol. 6, No. 6, 2011, pp. 1586-1591.

[2] J. P. O. Obiero, "Evaluation of Infiltration Using Green Ampt Model and Rainfall Runoff Data for Lagan and Sambret Catchments, Kericho, Kenya," M.Sc. Thesis, Department of Agricultural Engineering, University of Nairobi, Nairobi, 1996.

[3] A. Rasoulzadeh, "Estimating Hydraulic Conductivity Using Pedotransfer Functions," In: L. Elango, Ed., *Hydraulic Conductivity Issues, Determination, and Applications*, Intech, 2011, pp. 145-164. doi:10.5772/22753

[4] K. E. Saxton and W. J. Rawls, "Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions," *Soil Science Society of America Journal*, Vol. 70, No. 5, 2006, pp. 1569-1578.

[5] M. G. Schaap and F. J. Leij, "Using Neural Networks to Predict Soil Water Retention and Soil Hydraulic Conductivity," *Soil and Tillage Research*, Vol. 47, No. 1-2, 1998, pp. 37-42. doi:10.1016/S0167-1987(98)00070-1

[6] J. Tomasella and M. Hodnett, "Pedotransfer Functions in Tropical Soils," In: Y. Pachepsky and W. J. Rawls, Eds., *Development of Pedotransfer Functions in Soil Hydrology. Developments in Soil Science*, Vol. 30, Elsevier B. V., Amsterdam, 2004, pp. 415-429.

[7] H. Vereecken and M. Herbst, "Statistical Regression," In: Y. Pachepsky and W. J. Rawls, Eds., *Development of Pedotransfer Functions in Soil Hydrology. Developments in Soil Science*, Vol. 30, Elsevier B. V., Amsterdam, 2004, pp. 3-19.

[8] R. T. Walczack, F. Moreno, C. Slawinski, E. Fernandez and J. L. Arrue, "Modelling of Soil Water Retention Curve Using Soil Solid Phase Parameters," *Journal of Hydrology*, Vol. 329, No. 3-4, 2006, pp. 527-533. doi:10.1016/j.jhydrol.2006.03.005

[9] T. Wells, S. Fityus, D. W. Smith and H. Moe, "The Indirect Estimation of Saturated Hydraulic Conductivity of Soils, Using Measurements of Gas Permeability. I. Laboratory Testing with Dry Granular Soils," *Australian Journal of Soil Research*, CSIRO Publishing, 2006. http://www.highbeam.com

[10] J. H. Wosten, M. Y. A. Pachepsky and W. J. Rawls, "Pedotransfer Functions: Bridging the Gap between Available Basic Soil Data and Missing Soil Hydraulic Characteristics," *Journal of Hydrology*, Vol. 251, No. 3-4, 2001, pp. 123-150. doi:10.1016/S0022-1694(01)00464-4

[11] N. H. Batjes, Ed., "A Homogenized Soil Data File for Global Environmental Research: A Subset of FAO, ISRIC, and NRCS Profiles (Version 1.0)," Working Paper and Preprint 95/10b, International Soil Reference and Information Centre, Wageningen, 1995. http://www.isric.org/NR/exeres/545B0669-6743-402B-B 79A-DBF57E9FA67F.htm

[12] M. van Genutchen, "A Closed Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils," *Soil Science Society of America Journal*, Vol. 44, No. 5, 1980, pp. 892-898. doi:10.2136/sssaj1980.03615995004400050002x

Appendix

Analysis of Cross Correlations

Table A1. Cross correlation tests for transformed predictor variables for θ_s^{-1}.

	Sand	Silt^{-1}	BD^3	$\sqrt{\text{clay}}$	ln(orgC)
Sand	$r = 1$	$r = 0.42$	$r = 0.26$	$r = --0.71$	$r = -0.10$
Silt^{-1}	$r = 0.42$	$r = 1$	$r = 0.19$	$r = -0.13$	$r = -0.25$
BD^3	$r = 0.26$	$r = 0.19$	$r = 1$	$r = -0.28$	$r = -0.59$
$\sqrt{\text{clay}}$	$r = -0.71$	$r = -0.13$	$r = -0.28$	$r = 1$	$r = 0.08$
ln(orgC)	$r = -0.10$	$r = -0.25$	$r = -0.59$	$r = 0.08$	$r = 1$

Table A2. Cross correlation tests for transformed predictor variables in e^{θ_r}.

	Sand	Silt^{10}	BD^2	$\sqrt{\text{clay}}$	ln(orgC)
Sand	$r = 1$	$r = -0.25$	$r = 0.23$	$r = -0.71$	$r = -0.10$
Silt^{10}	$r = -0.25$	$r = 1$	$r = 0.06$	$r = -0.21$	$r = -0.09$
BD^2	$r = 0.23$	$r = 0.06$	$r = 1$	$r = -0.24$	$r = -0.09$
$\sqrt{\text{clay}}$	$r = -0.71$	$r = -0.21$	$r = -0.24$	$r = 1$	$r = 0.08$
ln(orgC)	$r = -0.10$	$r = -0.09$	$r = -0.62$	$r = 0.08$	$r = 1$

Table A3. Cross correlation tests for transformed predictor variables in $\sqrt{\alpha}$.

	e$^{\text{sand}}$	silt^{-2}	BD	clay^{-1}	ln(orgC)
e$^{\text{sand}}$	$r = 1$	$r = 0.89$	$r = 0.13$	$r = 0.3$	$r = -0.21$
silt^{-2}	$r = 0.89$	$r = 1$	$r = 0.14$	$r = 0.23$	$r = -0.24$
BD	$r = 0.13$	$r = 0.14$	$r = 1$	$r = 0.22$	$r = -0.64$
clay^{-1}	$r = 0.30$	$r = 0.23$	$r = 0.22$	$r = 1$	$r = 0.08$
ln(orgC)	$r = -0.21$	$r = -0.24$	$r = -0.64$	$r = 0.08$	$r = 1$

Table A4. Cross correlation tests for transformed predictor variables in $\sqrt{n}$.

	sand^5	Silt^{-1}	ln(clay)	orgC^{-1}
sand^5	$r = 1$	$r = 0.74$	$r = -0.69$	$r = 0.34$
Silt^{-1}	$r = 0.74$	$r = 1$	$r = -0.25$	$r = 0.52$
ln(clay)	$r = -0.69$	$r = -0.25$	$r = 1$	$r = -0.21$
orgC^{-1}	$r = 0.34$	$r = 0.52$	$r = -0.21$	$r = 1$