Dryland Soil Research, Vol. 2, No. 1, pp. 65-79, 2025
Type of article: Original, Available online at www.dlsr.iut.ac.ir.
https://doi.org/10.47176/dlsr.02.01.1047

A data-driven approach to predict soil hydraulic conductivity: GMDH compared with ANN and multiple regression

Mehdi Rahmati[1,2†]

[1] *Department of Soil Science and Engineering, Faculty of Agriculture, University of Maragheh, Maragheh, Iran,*
[2] *Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences: Agrosphere (IBG-3), Jülich, Germany*

†*Corresponding Author Email: mehdirmti@gmail.com*

(Received 27 September 2025, Revised 06 December 2025, Accepted 08 December 2025)

ABSTRACT

Accurate estimation of saturated hydraulic conductivity (K_s) is essential for soil and water management, yet the reliability of the pedotransfer functions (PTFs) is often overlooked.Tthis study compares the predictive performance and the robustness of field estimates of K_s obtained by three types of PTFs: Multiple Regression (MR), Artificial Neural Networks (ANN) and the Group Method of Data Handling (GMDH) developed using 134 soil samples collected in north-western Iran, which is a region with semi-arid conditions and mixed agricultural uses whereas the dataset encompasses a wide range of structural and textural variationsto K_s prediction. In addition to traditional soil properties soil moisture deficit compared to the optimum value at the time of sampling, θ_d, was used as a proxy indicator of soil structural condition. Model precision was evaluated using Root Mean Square Error (RMSE) and the Nash–Sutcliffe efficiency; reliability was determined through repeated data splitting. Even though ANN provided good accuracy for the training set, its performance for the validation set was inconsistence. MR produced consistent, albeit limited performances over both the training and validation subsets. Conversely, GMDH appears to strike a good compromise between prediction accuracy, reliability, parsimony of the predictor set, and texture versus structure variables. The results point to the importance of including structural measures such as θ_d in PTF development and provide a basis for considering model repeatability a long with accuracy. In general, the results indicate that GMDH is a robust and feasible technique to develop accurate PTFs for K_s predictions with limited amount of data.

Keywords: Pedo-transfer function, soil function modeling, Water retention curve, Artificial neural networks, multiple regression.

1. Introduction

Characterization of soil hydraulic properties is required to describe the water, energy, and carbon exchange processes between the land surface and the atmosphere (Montzka et al., 2017) and to assess soil and ground water contamination risk or soil remediation activities (Neyshaboury et al., 2015). The soil water retention curve, WRC, the hydraulic conductivity curve, HCC, and soil water diffusivity, D(θ), are the most important and fundamental hydraulic properties of soil (Montzka et al., 2017; Rahmati and Neyshaboury, 2016). Vereecken et al. (2016) pointed out that the relative magnitude of soil water fluxes is collectively controlled by WRC, HCC and D(θ).

Saturated hydraulic conductivity (K_s) is one of the soil hydraulic properties usually required as input in simulation models (Alvarez-Acosta et al., 2012; Herbst et al., 2006) to derive the HCC. K_s is the maximum flow rate of water under saturated condition (Mallants et al., 1997b) which is used to assess the risk of leaching of solutes, to characterize water infiltration, and to model surface runoff (Masís-Meléndez et al., 2014). Further, it is widely used as a scaling factor to describe the unsaturated hydraulic conductivity (Brooks and Corey, 1966; Campbell, 1974; Doussan and Ruy, 2009; Kosugi, 1996; Mualem, 1976; Neyshabouri et al., 2013, 2015; Van Genuchten, 1980). The direct measurements of K_s in the laboratory or in the field (e.g., according to Klute and Dirksen, 1986; Reynolds and Elrick, 1985; Reynolds et al., 2002) is usually costly, time consuming, and tedious (Alvarez-Acosta et al., 2012; Christiaens and Feyen, 2001; Islam et al., 2006; Jabro, 1992; Montzka et al., 2017; Schaap and Leij, 1998; Tietje and Hennings, 1996) and subject to large uncertainty due to associated small scale soil heterogeneity and experimental errors (Aimrun and Amin, 2009; Alvarez-Acosta et al., 2012; Schaap and Leij, 1998). In fact, several studies showed that the relative accuracy (expressed as a percentage) of different methods varies among different soil types (Gupta et al., 1993; Mallants et al., 1997a; Mohanty et al., 1994; Paige and Hillel, 1993). On the other hand, the global application of land surface models requires knowledge of K_s while global measurements of K_s are not feasible, and data are not available to provide global coverage (Vereecken et al., 2016). Therefore, soil scientists, hydrologists, or environmentalists usually attempt to estimate K_s using the easiest and fastest approaches where pedo-transfer functions (PTFs) are among the most applied methods

(Pachepsky and Rawls, 1999). PTFs allow translating soil information that is available to information that is needed or required in e.g., simulation models (Bouma, 1989). Researchers usually apply PFTs to fill the gap between commonly available soil properties and those soil characteristics, which are required as inputs to various types of models (McBratney et al., 2002). Most of the PTFs related to the estimation of the K_s are based on the following exponential form (Brakensiek et al., 1984; Campbell and Shiozawa, 1992; Cosby et al., 1984; Dane and Puckett, 1994; Julia et al., 2004; Puckett et al., 1985; Saxton et al., 1986; Vereecken et al., 1990; Wösten, 1997; Wösten et al., 1999):

$$K_s = \text{constant} \times \exp(f(x)) \quad [1]$$

where, several soil properties serve as explanatory variable *f(x)* such as clay, silt, and sand content, organic carbon (OC) or organic matter (OM) content, bulk density, and porosity or saturated water content (Brakensiek et al., 1984; Campbell and Shiozawa, 1992; Cosby et al., 1984; Dane and Puckett, 1994; Puckett et al., 1985; Saxton et al., 1986; Vereecken et al., 1990; Wösten, 1997; Wösten et al., 1999). Additionally, several other linear or nonlinear regression based PTFs have been also developed for K_s prediction, e.g., by Jabro (1992), Suleiman and Ritchie (2001), Julia et al. (2004), and Spychalski et al. (2007). Artificial neural network (ANN) is also widely applied to derive PTFs to predict K_s. Several researchers, e.g., Minasny et al. (2004), Merdun et al. (2006), Parasuraman et al. (2006), Agyare et al. (2007), Ghanbarian-Alavijeh et al. (2010), Arshad et al. (2013), Albalasmeh et al. (2022), Yamaç et al. (2022), Mozaffari et al. (2024), Moosavi et al. (2024), Naderianfar (2025), Mozaffari et al. (2025), and Elbisy (2025), used ANN to develop PTFs for K_s prediction. Schaap et al. (1998), Arshad et al. (2010), and Sarmadian and Taghizadeh-Mehrjardi (2014) reported that ANN produced more accurate PTFs to predict soil properties than regression techniques. On the other hand, other studies confirm higher efficiency for regression-based PTFs compared to ANN-based ones. For example, Zhao et al. (2016) reported that although ANN and MR showed similar accuracy in terms of K_s prediction, the MR-based PTF showed higher reliability compared to the ANN-based PTFs.

The numerous classical PTFs established over the past few decades can in general be divided into two groups: empirical and physically based approaches. Empirical PTFs (from simple linear and nonlinear regressions to more flexible machine-learning approaches) aim for statistical relationships between easily measured soil properties, and hydraulic parameters like K_s. They are powerful due to their simplicity and low data need although they tend to heavily rely on the representativeness of calibration data and the lack of extrapolation in unknown scenarios remains questioning. In contrast, physically based or semi-empirical PTFs try to integrate mechanistic knowledge on soil water flow, by means of a functional relation between K_s and pore size distribution, soil structure or water retention parameters. Although they are mathematically attractive, these models rely on structural or hydraulic parameters that are seldom measured at regional or larger scales. Consequently, despite their drawbacks, many practical applications are still based on empirical PTFs. This dichotomy brings to focus one of the central challenges posed by these two types of models: empirical PTFs might provide good accuracy but often show no interpretable relationship with soil processes, while mechanistic/similarity-based approaches deliver physical insight but require that we have data which are prohibitively costly to collect. Hence, there is a requirement for approaches that can achieve the compromise between predictive performance and the selection of those with the highest impact on soil properties, where GMDH could have an advantage over traditional empirical models.

Aside from applied procedure, the number of applied input variables is also deterministic in PTFs application. For example, the higher accuracy of the predictions by ANN will be accessible by including as many as number of predictors. While the higher the number of predictors, the higher the cost for data collecting. Therefore, regarding the practical point of view, it is important to identify the most effective predictors for final PTF development. Literature review reveals that soil clay, silt, and sand content beside its organic carbon content, bulk density, and dryness and wetness condition are among the most applied properties for PTFs development (Brakensiek et al., 1984; Campbell and Shiozawa, 1992; Cosby et al., 1984; Dane and Puckett, 1994; Puckett et al., 1985; Saxton et al., 1986; Vereecken et al., 1990; Wösten, 1997; Wösten et al., 1999). However, the accuracy of these empirical PTFs outside the database used for their development is unknown (Vereecken et al., 2016). Therefore, there seems to be no guaranty if these communally used properties will serve as proper indicators for K_s prediction in all conditions. On the other hand, it is always argued that a good estimate of K_s cannot be obtained without including structural information (Or et al., 2021; Vereecken et al., 2022; Sharghi et al., 2025). However, Logsdon et al. (2013) pointed out that there is only an extremely limited quantitative understanding of soil structure and dynamics and how they affect various functions of soil. So far, several indicators including soil aggregate stability, aggregates mean diameter, pore size distribution, and fractal geometry (Grossman et al., 2002) have been introduced to quantify soil structure. However, the variability of the soil structure is a crucial factor which is missed in our investigations. More importantly, when soil is sampled for water movement characterization, it is urgent to take as many representative samples as possible with the minimum change in soil structure. It is strongly suggested to obtain undisturbed soil samples at/or near field capacity (Grossman et al., 2002) to avoid soil

shattering or compaction during sampling to prevent soil structure damages. However, in most cases, especially in arid or semi-arid regions, the desirable soil water content for soil sampling is not present and soil is usually sampled beyond its optimum water content. Therefore, it is necessary to provide a proper indicator to quantify the uncertainties in K_s measurement which may be relevant to the variability of the soil structure due to soil sampling in non-optimal conditions. Therefore, we further introduce applying soil moisture deficit from its optimum value for sampling (θ_d) as an indicator of soil structural variability. Soil moisture deviation from its optimum value is expected to affect the stability of soil structure and may thus also affect K_s. Therefore, soil water content deviation from its optimum value may reflect its effect on K_s measurement.

As stated above, the accuracy of the developed PTFs does indeed depend on the choice of the procedure (Weihermüller et al., 2021). Developing more dependable PTFs, determining the most effective input variables, and identifying soil groups that could improve PTFs accuracy are as important as their accuracies. Nevertheless, a review of literature reveals that most K_s-related PTFs were developed without a reliability test (at least up to our knowledge). On the other hand, although stepwise regression can be applied to identify the set of the most effective predictors, regression based PTFs usually expose a lower accuracy compared to ANN predictions (Arshad et al., 2010; Sarmadian and Taghizadeh-Mehrjardi, 2014; Schaap et al., 1998). However, ANN does not provide an explicit procedure to select the most essential (statistically relevant) PTFs input variables (Pachepsky et al., 1996). In contrast to ANN, the group method of data handling (GMDH), which finds an approximate relationship between a set of input and output variables (Farlow, 1984; Pachepsky et al., 1998), enables the identification of essential input variables (Rahmati, 2017). In fact, the GMDH retains only essential input variables in a flexible net of regression equations due to a built-in algorithm (Pachepsky and Rawls, 1999). Hecht-Nielsen (1990) showed that GMDH is more appropriate than statistical regression to link the substantial number of variables in a complicated relationship between input and output variables. Yet, no comparison has been made between GMDH and two other commonly used procedures (ANN and MR) to evaluate their performance in terms of a reliable and accurate estimation of K_s.

Despite many efforts to derive PTFs for the prediction of K_s, there are still two major gaps. First, most research uses MR or ANN even though these methods do not have an explicit device for choosing predictive predictors (ANN) and lose robustness when many predictors are used in the calculations (MR). Therefore, the comparative evaluation of the performance of GMDH—a method whose structure has been developed precisely for identifying the most influential input parameters—has not yet been systematically reported in K_s prediction. Second, while soil structure is generally recognized as essential for estimating K_s, the structural variability at sampling (e.g., by deviations from optimal moisture) is seldom formally quantified or included in the development of PTF. Thereby, the observed PTFs do not tell us much about how soil structural attributes affect model predictive accuracy and reliability. To bridge these gaps, we need to compare GMDH with ANN and MR under the framework that explicitly includes soil texture and structure variables. Therefore, the objectives of this study were: 1) evaluate the prediction accuracy and reliability of GMDH, ANN and MR for K_s; 2) employ GMDH to select the most important input variables; 3) determine whether soil moisture deficit at sampling (θ_d) helps for better K_s estimation; and 4) compare the resulting PTFs to the more frequently used exponential model (Eq. 1).

2. Methods and Materials

2.1. Soil Sampling and Field/Laboratory Measurements

Disturbed and undisturbed soil samples from 0- 15 cm depths were collected at 134 distinct locations in the North-western Iran, located between latitudes 37°43′07" N to 37°50′08" N and longitudes 46°22′23" E to 46°28′05" E. According to Rahmati et al (2020), the watershed is characterized by an area of about 7,854 hectares, spreading across various elevations ranging from around 3,534m at the high lands to around 2,190 m at the outlet point of the watershed (Figure 1), with annual average precipitation rate amounting to 320 mm. Bare-lands (46%) and poor pastures (42%) are the dominant land uses of the study area, which is only 12% farming that included both rainfed and irrigated fields (Figure 1) (Rahmati et al., 2015).

Disturbed soil samples were analyzed for clay (*cc*), silt (*si*), and sand (*sa*) contents using the hydrometer method (Gee and Or, 2002), organic carbon (*OC*) using wet oxidation technique (Nelson and Sommers, 1982), aggregate stability (*WAS*) using wet-sieving method (Nimmo and Perkins, 2002), particle density (D_p) by pycnometry (Flint and Flint, 2002), and electrical conductivity (*EC*) in paste saturation extracts by EC-meter. We used undisturbed core samples to determine saturated hydraulic conductivity (K_s) by the falling head method (Reynolds et al., 2002), and bulk density (D_b) according to the method proposed by Grossman and Reinsch (2002). Corresponding soil water content at sampling time (θ_i) and field saturated water content (θ_{fs}) were measured using a gravimetric method. Since measured field capacity (FC) was not available, we used the equation proposed by Saxton et al. (1986) as a best guess for FC and thus to derive the soil moisture deficit from its optimum value (θ_d) at sampling time (θ_d=0.9×FC-θ_i). We do recognize, however, that the use of Saxton's equation includes a model-dependent approximation for FC and hence θ_d has added uncertainty based on this PTFs estimate rather than having been directly observed.

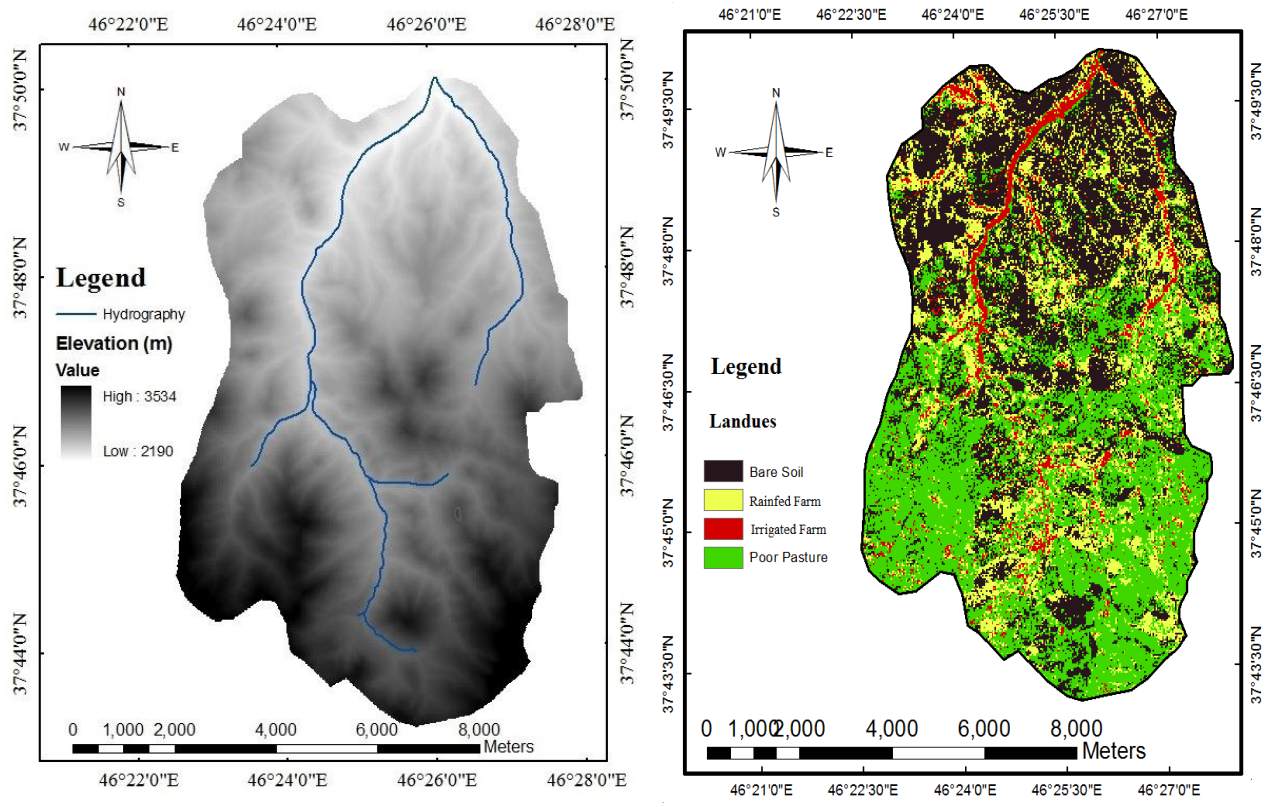

Figure 1. DEM (left) and land use (right) maps of the studied area (Lighvan Watershed), in northwest of Iran (Rahmati et al., 2020)

Table 1. Statistical parameters of measured characteristics in the current study

Parameters		Maximum	Minimum	Mean	CV (%)
Texture	Clay (-)	0.352	0.067	0.169	31.95
	Silt (-)	0.480	0.070	0.271	26.70
	Sand (-)	0.795	0.342	0.560	17.21
Aggregate stability (%)		95.92	25.14	65.39	28.63
Saturated hydraulic conductivity (cm/h)		18.39	1.032	6.370	60.66
Bulk density (g/cm^3)		1.486	1.207	1.345	3.81
Particle density (g/cm^3)		2.646	2.212	2.481	4.29
Organic carbon (%)		2.048	0.098	0.867	48.92
Electrical conductivity (mS/cm)		1.200	0.300	0.702	36.57
Antecedent water content (cm^3/cm^3)		0.140	0.102	0.124	5.90
Field saturated water content (cm^3/cm^3)		0.575	0.416	0.505	7.09

CV: Coefficient of variation

The sampled soils cover four textural classes including sandy loam, sandy clay loam, loam, and clay loam (Figure 2). Table 1 summarizes the statistical measures of soil properties where K_s had the highest variation with a coefficient of variation of 61 % ranging from 1.032 to 18.30 cm/h. Bulk and particle densities with a coefficient of variation of about 4 % showed the lowest variation.

2.2. PTFs development

2.2.1. Multiple regression

A simple linear multiple regression (MR) was applied to predict K_s from the other measured soil characteristics:

$$K_s=a+b_1X_1+b_2X_2+...+b_iX_i \quad [2]$$

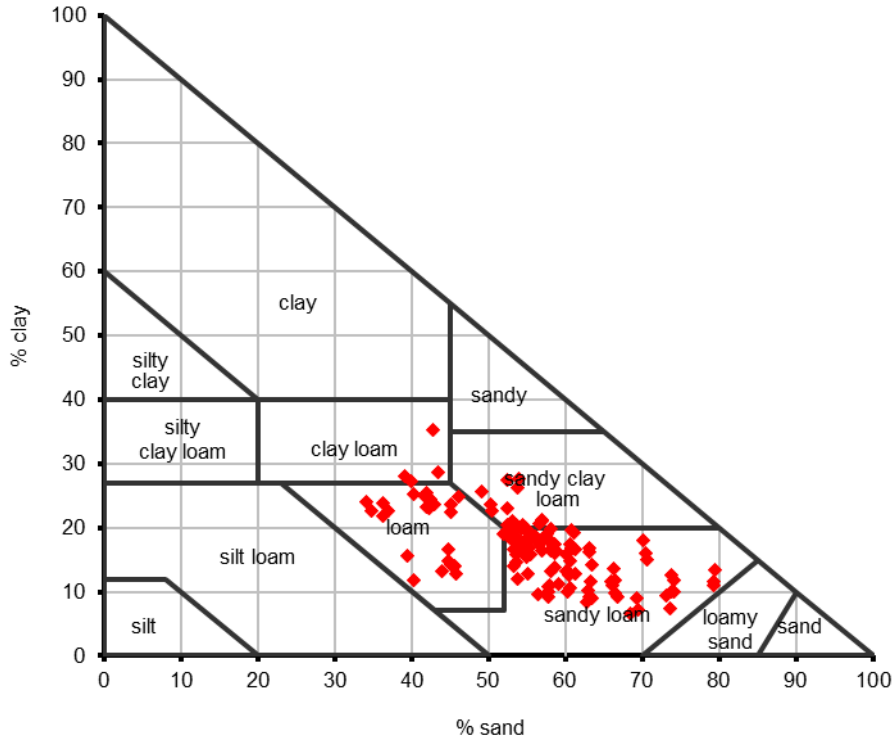

Figure 2. Distribution of investigated soils among different soil texture classes

where, α is the intercept, b_1 to b_i are regression coefficients, and X_1 to X_i implies the soil characteristics.

2.2.2. Artificial neural networks

Feedforward artificial neural network (ANN) was applied to develop a non-linear PTF for the estimation of K_s. Naming input and output variables and normalizing the data between 0 and 1 is the first step in developing ANN models. In the second step, we figured out the best ANN architecture by applying a trial-and-error method to find the optimal number of hidden neurons through training of various architectures. Once the best ANN architecture is trained, we applied it to independent data to confirm it. For the ANN network, the best architecture consists of ten neurons in the input layer, twenty-four neurons in the hidden layer (found by user-defined GridSearchCV method), and one neuron in output layer with trainscg and purelin threshold functions for hidden and output layers, respectively. The trainscg is a network training function that updates weight and bias values according to the scaled conjugate gradient method (Møller, 1993). While purelin is a linear transfer function to calculate a layer's output from its net input.

It is worth noting that we intentionally retained a simple ANN architecture to fall in line with most common PTF applications: The goal of our work was the comparison between methods and not perform extensive tuning or regularization of a single model.

2.2.3. Group method of data handling

We applied group method of data handling (GMDH) (Pachepsky and Rawls, 1999) to predict K_s using the measured soil characteristics as input variables. The following quadratic regression was used to obtain the preliminary estimates (z_{ij}) for the first layer of the GMDH network:

$$z_{ij} = c_0 + c_1 x_i + c_2 x_j + c_3 x_i^2 + c_4 x_j^2 + c_5 x_i x_j \quad [3]$$

where x_i and x_j are pairwise selections of input variables and c_0 to c_5 are the polynomial coefficients. The total number (n) of polynomials is decided by following equation:

$$n = \frac{N \times (N-1)}{2} \quad [4]$$

where N is the number of input variables. To develop the GMDH network, first, all polynomials were found using pairwise selected variables (x_i and x_j). Then, the least effective new variables were screened out using the following criterion (Farlow, 1984):

$$e = p \times RMSE_{lowest} + (1-p) \times RMSE_{highest} \quad [5]$$

where e is the indicator used to select new variables and p is the selection pressure implying a number between 0 and 1. Higher numbers show higher pressure in the new variable selection. The $RMSE_{lowest}$ and $RMSE_{highest}$ respectively are the lowest and highest root-mean-square-errors between target and preliminary estimates. We set the p values equal to 0.75 applying a try and error method

to optimize the network. The preliminary estimates with a root mean square error (RMSE) lower than *e* were selected for the next layer. The polynomials then were further improved by repeating steps 1 and 2 and using new selected variables (z_{ij}'s) from the earlier step till the smallest value of the selection criterion obtained from the current iteration shows no improvement in relation to the smallest value obtained from the previous iteration (Pachepsky and Rawls, 1999). The version of the GMDH algorithm used in this study is coded in Matlab.

2.2.4. Data preparation and division into train and test subsets

To develop and validate the PTFs, we randomly split the data into two datasets, one training dataset and an independent validation subset making up 30% of total data. Prior to any modeling, we normalized both original input and output variables to have zero mean and unit variance and employed the normalized variables in the PTFs development using the following equation:

$$Z_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)} \quad [6]$$

where, X_i and Z_i represent measured and normalized data, respectively.

2.3. Statistical Analysis

The terms accuracy and reliability are two measures that will be used in this work to assess the quality of the different methods. The term accuracy technically means the degree to which the result of a measurement, calculation, or specification conforms to the correct value or a standard (Webster, 2006). While the term reliability technically means the degree to which the result of a measurement, calculation, or specification can be depended on to be exact (Webster, 2006). More simply, we consider the term accuracy as a measure which evaluates the results to be close to reality and term reliability as a measure which evaluates the results to be repeatable.

2.3.1. PTFs accuracy assessment

We evaluated the PTF accuracy by applying the root mean square error (RMSE) and the Nash-Sutcliffe coefficient (*E*) (Nash and Sutcliffe, 1970):

$$RMSE = \sqrt{\frac{\sum (LogX_m - LogX_p)^2}{n}} \quad [7]$$

$$E = 1 - \frac{\sum (X_m - X_p)^2}{\sum (X_m - \bar{X}_m)^2} \quad [8]$$

where X_m and X_p are measured and predicted K_s, respectively. The $\bar{X}_m$ is the mean value of measured K_s. A value of RMSE close to zero shows a high accuracy of the method. The *E* varies between $-\infty$ and 1, where the later shows perfect match. The $E < 0$ shows that the model performs worse than simply using the mean of the observed values.

2.3.2. PTFs reliability assessment

In order to evaluate the reliability, or repeatability, of the models, the random splitting of data into train and independent subsets was repeated 10 times and RMSE and *E* values were calculated for each replication, and average RMSE and *E* values, along with their variances, were derived from the replications for both train and independent subsets (Pachepsky and Rawls, 1999). Then, the following function was applied to compute the reliability of each method:

$$Reliability(\%) = 100 - CV(\%) \quad [9]$$

where, CV is the coefficient of variation of RMSE or *E* among the ten replicates. A reliability value of one hundred means that the applied method resulted in the same accuracy among various replicates while the reliability values lower than one hundred means that the accuracy of the applied methodology is variable.

It should be noted that the repeated random split approach used in this paper is a special case of cross-validation known as repeated random sub-sampling or Monte Carlo cross-validation. It has been proposed and suggested for analysis of PTFs with small number of observations where the main concern is to evaluate prediction accuracies as well as stability of the model (Pachepsky and Rawls, 1999). Ten repetitions of the train–test partition yield a stable estimate of model variance and obviates the need for further regularization strategies that are usually necessary only when overfitting cannot be assessed by means of replicates. So that the CV among replicates is a direct measure of how reliable (repeatable) each modelling technique is.

2.3.3. Statistical comparisons among different methodologies

To conduct a statistical comparison among the methods, a *t* test using pooled estimates of the mean square error (MSE_p) was conducted (Sirkin, 2006):

$$t = \frac{\left| x_i - x_j \right|}{\sqrt{MSE_p \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \quad [10]$$

where *X* implies the criterion applied to measure PTF accuracy, *i* and *j* are pairwise selections of the applied methods, and n_1 and n_2 are the number of observations used for the PTF development of each selected method.

Table 2. The summary of statistical analysis of train and test subsets for ten replications of MR methodology for K_s prediction

Variable	Train		Train	
	E	RMSE	E	RMSE
Mean	0.710	0.118	0.654	0.124
Min	0.674	0.111	0.365	0.113
Max	0.729	0.121	0.754	0.148
CV (%)	2.36	2.67	17.04	9.11
Reliability (%)	97.64	97.33	82.96	90.89

CV: Coefficient of variation

Table 3. The summary of statistical analysis of train and test subsets for ten replications of ANN methodology for K_s prediction

Variable	Train		Train	
	E	RMSE	E	RMSE
Mean	0.790	0.097	0.371	0.162
Min	0.465	0.058	-0.464	0.113
Max	0.939	0.163	0.682	0.230
CV (%)	16.86	30.28	98.71	18.46
Reliability (%)	85.14	69.72	1.29	81.54

CV: Coefficient of variation

The parameter MSE_p was calculated according to:

$$\mathrm{MSE_p} = \frac{(\mathrm{n_i}-1)\mathrm{MSE_i} + (\mathrm{n_j}-1)\mathrm{MSE_j}}{(\mathrm{n_i}-1)+(\mathrm{n_j}-1)} \quad [11]$$

where n implies the number of observations and MSE is the mean square error between measured and predicted values of K_s and subscripts i and j show the two methods compared.

3. Results and Discussion

3.1. PTF accuracy and reliability

3.1.1. MR procedure

Table 2 reports the summary of statistical analysis of train and test subsets for MR for K_s prediction. In our analysis we also included linear interactions in terms of soil properties and K_s.

PTFs derived from the training set showed a high accuracy with a mean E and RMSE of 0.71, and 0.118, respectively (Table 2). Application of the PTFs on the validation dataset resulted in a mean E and RMSE of 0.65, and 0.124, respectively. Reliability was also given as shown by values of 98 and 97 % for train dataset and 83 to 91 % for the validation datasets.

Since PTFs development was repeated ten times, only the last one is reported here. Therefore, the following PTF was developed to predict K_s at 10th replication showing E and RMSE of 0.70 and 0.121, respectively, for the train subsets and E and RMSE of 0.75 and 0.113 for the validation subset.

$$K_s = 7.18 - 4.20cc - 5.46si - 6.05sa - 0.21D_b + 0.10OC - 0.19D_p - 0.27\theta_{fs} - 0.58\theta_d + 0.01EC + 0.04WAS \quad [12]$$

where, all input and output variables represent normalized values where D_b and D_p are bulk and particle densities in g/cm^3, θ_{fs} is field saturated water content in volumetric percent, θ_d is soil moisture deficit from its optimum value for sampling in volumetric percent, WAS, cc, si, sa, and OC are wet-aggregate stability, clay, silt, sand, and organic carbon contents in percent and EC is electrical conductivity in dS/m. Figure 3 shows a scatter plot of measured and predicted K_s using Eq. 12 on the complete dataset.

3.1.2. ANN procedure

A feedforward ANN model was applied to predict K_s. Like MR modeling, all variables were normalized to achieve an effective training of the network (Luk et al., 2000). However, the effect of normalization diminishes as network and sample size become larger (Luk et al., 2000).

A three-layered feed forward architecture with one input layer, one hidden layer, and one output layer was developed to predict K_s. Table 3 reports the statistical analysis of ten replications between measured and ANN–simulated values of K_s for the train and validation subsets. The mean E and RMSE between measured and predicted K_s for the train data set were 0.79 and 0.097 showing high accuracy. However, applying the validation dataset as an independent data to check the model's accuracy depicted low accuracy showing a mean E and RMSE of 0.37 and 0.162. The ANN showed low reliability values of 70 to 85 % for the train dataset and 1 to 82 % for the validation datasets. Although, the CV for E and RMSE was comparably low for the train subset (< 30%), a high CV up to 100 % for the validation subset showed low reliability. Figure 4 shows scatter plots of measured and predicted K_s using the ANN network applied to the train dataset and the validation dataset.

It is worth noting that the large gap between E values for training (0.79) and testing (0.37) revealed overfitting of the ANN model. This is not surprising due to the limited size of the data, and the ability of ANN structures that can fit in a fine manner to training examples but generalize badly when sample size is small. While methods including early stopping, regularization, dropout or data augmentation can reduce overfitting, optimizing the ANN architecture is outside of the scope of the current work that focuses on model class comparisons under similar

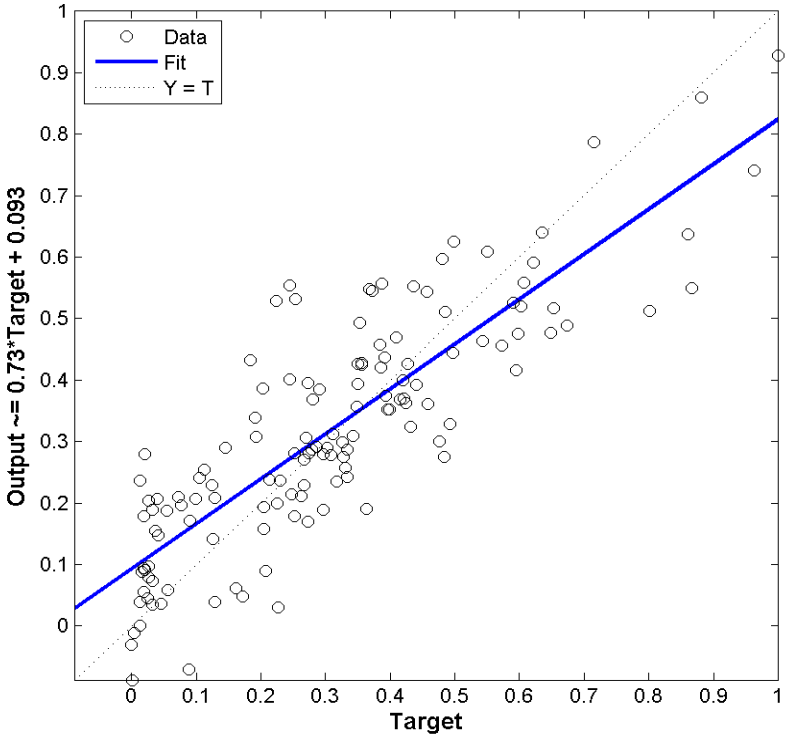

Figure 3. Scattering measured (Target) and predicted (Output) values of normalized K_s applying Eq. 12 to the full dataset, 1:1 line is dashed.

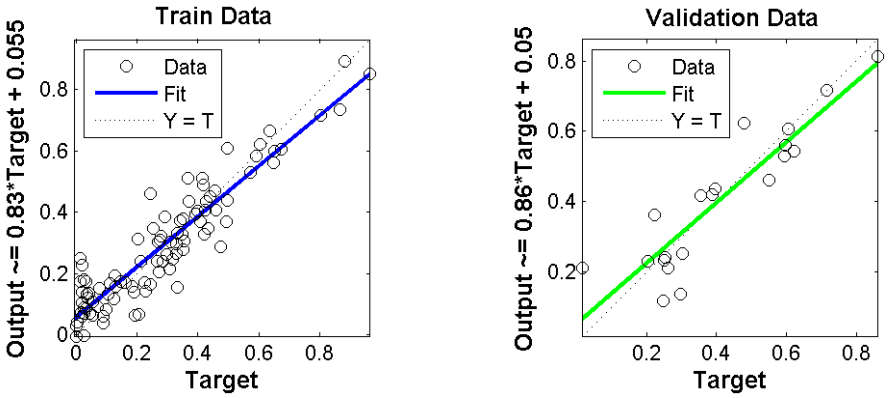

Figure 4. Scattering measured (Target) and predicted (Output) values of normalized K_s applying ANN network to the train dataset and the validation in 10th replicate.

conditions rather than optimization for individual models. The observed overfitting additionally illustrates the difficulty of using ANN-based PTFs for moderate-sized datasets and indicates the relatively consistent performance exhibited by GMDH.

3.1.3. GMDH procedure

To develop GMDH network, all data sets were also normalized to vary between 0 and 1. In order to prevent the network being too complicated to interpret and in order to check the different networks with different number of layers and neurons, we repeated the network development applying different numbers of layers (2, 3, and 5) and neurons (5, 10, and 15). Then a comparison was carried out to select the best network. The network based on a lower number of layers and neurons and with better accuracy and reliability was preferred. Table 4 reports the summary of the statistical analysis of the different GMDH network architectures. Results revealed that increasing the number of the GMDH network not only showed not much difference in network accuracy but also resulted in a low accuracy in some cases (Table 4). Therefore, the network

Table 4. The summary of statistical analysis of different GMDH network architectures for K_s prediction

GMDH network architecture	Train subset		Test subset	
	E	RMSE	E	RMSE
Two layers and five neurons	0.627	0.133	0.699	0.124
Two layers and ten neurons	0.626	0.136	0.612	0.132
Two layers and fifteen neurons	0.649	0.130	0.591	0.140
Three layers and five neurons	0.633	0.136	0.594	0.130
Three layers and ten neurons	0.654	0.129	0.690	0.121
Three layers and fifteen neurons	0.695	0.120	0.693	0.122
Five layers and five neurons	0.638	0.134	0.646	0.124
Five layers and ten neurons	0.673	0.128	0.644	0.123
Five layers and fifteen neurons	0.631	0.132	0.572	0.135

Table 5. The summary of statistical analysis of train and test subsets for ten replications of GMDH methodology (2 layers and five neurons) for K_s prediction

Variable	Train		Train	
	E	RMSE	E	RMSE
Mean	0.627	0.133	0.699	0.124
Min	0.558	0.118	0.619	0.099
Max	0.695	0.147	0.795	0.137
CV (%)	7.51	6.37	7.69	9.52
Reliability (%)	92.49	93.63	90.48	94.24

CV: Coefficient of variation

with two layers and five neurons was selected for further assessment. The statistical analysis for the ten replications between measured and selected GMDH –estimated values of K_s for the train and validation subsets are reported in Table 5. The mean E and RMSE between measured and predicted K_s for the train dataset were 0.63 and 0.133, respectively showing high accuracy. According to the validation dataset, the results also showed that the model accuracy was comparable to the train dataset showing a mean E and RMSE of 0.70 and 0.124, respectively. The results also revealed reliability values higher than 90% for both the train and the validation subsets.

Equation 13 stands for the network developed to predict K_s in the 10th replication showing an E and RMSE of 0.66 and 0.129 for the train subset and 0.80 and 0.099 for the validation subset, respectively.

$$K_s = -0.14+0.86z_1+0.65z_2+0.52z_1^2 +0.54z_2^2-1.30z_1 z_2 \quad [13]$$

where, z_1 and z_2 are preliminary estimates of K_s which were calculated using following equations:

$$z_1=0.91-1.21D_b-1.63\theta_d+0.21D_b^2 +1.57\theta_d^2+1.52D_b\times\theta_d \quad [14]$$

$$z_2=0.18+0.04cc+0.31\theta_{fs}-0.38cc^2 +0.26\theta_{fs}^2-0.47cc\times\theta_{fs} \quad [15]$$

where, D_b is bulk density in g/cm^3, θ_d is soil moisture deficit from its best value at sampling time in volumetric percent, cc is clay in percent, and θ_{fs} is field saturated moisture content in volumetric percent. We need to note that like the MR-based PTF, all input variables need to be normalized prior to use and output is normalized, as well. Figure 5 shows scatter plots of measured and predicted K_s using the finally selected GMDH network (Eq. 13) applied to the full dataset in 10th replication.

3.2. PTFs comparison

A statistical comparison of the applied methodologies using a t test is reported in Table 6. Pairwise comparison of the three applied methodologies revealed that ANN resulted in higher accuracy in training stage where E (0.79 *vs*. 0.71 and 0.63) criterion was significantly higher ($P < 0.01$) than those of MR and GMDH (Table 6). On the other hand, MR resulted in higher accuracy compared to GMDH

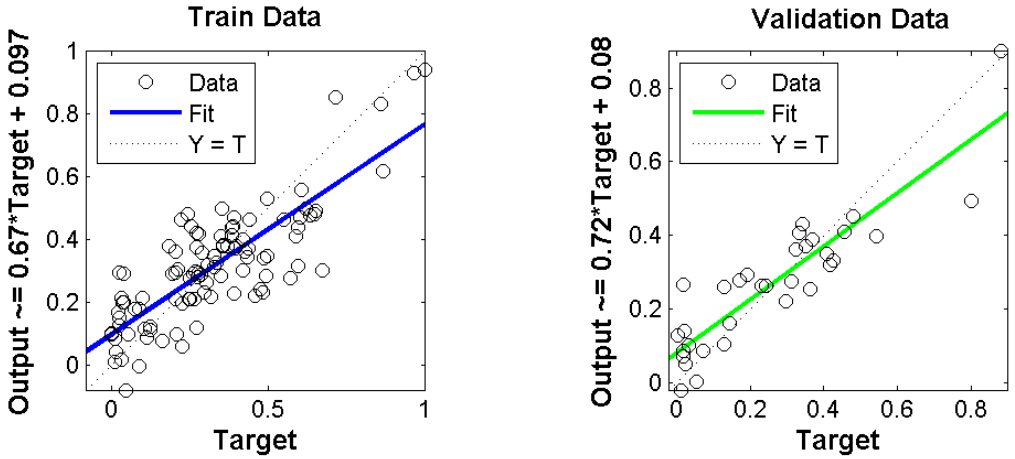

Figure 5. Scattering measured (Target) and predicted (Output) values of normalized K_s applying GMDH network and train and test data set in 10th replicate.

Table 6. The statistical comparisons of the applied methods using t test.

Subset	Variable	n	E_1 - E_2	MSE_1- MSE_2	MSE_P	t_{R^2}	t_E
Train	MR - ANN	94	0.71 - 0.79	0.014 - 0.009	0.012	5.78**	5.08**
	MR - GMDH	94	0.71 - 0.63	0.014 - 0.018	0.016	4.31**	4.53**
	ANN - GMDH	94	0.79 - 0.63	0.009 - 0.018	0.014	10.01**	9.60**
Test	MR - ANN	40	0.65 -0.37	0.015 - 0.026	0.021	4.15**	8.77**
	MR - GMDH	40	0.65 -0.70	0.015 - 0.015	0.015	0.43ns	1.62ns
	ANN - GMDH	40	0.37 -0.70	0.026 - 0.015	0.021	4.53**	10.17**

ns: insignificant and **: significant with P < 0.01

in training stage as the *E* of 0.71 for the MR method were significantly higher (P < 0.01) than that of GMDH procedure (*E* = 0.63). In contrast to the results on the train dataset, the results for the independent validation dataset revealed that GMDH resulted in better conformity between measured and predicted K_s where the *E* (0.70 vs. 0.65 and 0.37) was higher than MR (insignificantly) and ANN (significantly with P < 0.01) (Table 6). In addition to the better performance for validation dataset, the GMDH approach showed more reliability than the PTFs developed by MR and ANN. The CV of the *E* criterion of 7.7% was low for the GMDH showing more reliability, while the respective CV for MR and ANN were seventeen up to 100% (Figure 6).

The comparison between MR and ANN shows that although ANN provided higher accuracy for the train dataset, MR predicted K_s more accurately than ANN for the independent validation dataset. MR also resulted in more reliable PTF estimations than ANN showing a CV of 2 to 17% (for both train and validation datasets), whereas the respective CV varied between 14 to 99 % for the ANN approach. Regarding the accuracy term in train dataset, our results are in line with the results from Arshad et al. (2010); Arshad et al. (2013); Schaap et al. (1998), and Sarmadian and Taghizadeh-Mehrjardi (2014) reporting a higher accuracy for ANN compared to MR. However, our results revealed that the ANN fails in independent evaluation dataset compared to MR. Even in the case of higher accuracy for ANN in independent evaluation dataset, we showed here that the accuracy evaluation alone maybe is not sufficient to judge between several PTFs and reliability test joint with accuracy evaluation may present better measure in this regard.

In addition to producing more accurate and reliable PTFs, GMDH algorithm can help to determine the most important/effective input variables, which is another advantage of GMDH (Pachepsky and Rawls, 1999). We do believe that this may also be applicable through stepwise regression procedure. However, the reliability of PTFs is less for MR compared to GMDH. The GMDH approach comprised only four input variables including *cc*, D_b, θ_d, and θ_{fs} and produced more accurate and dependable PTFs, while the MR and ANN approaches included all ten input variables and resulted in lower or

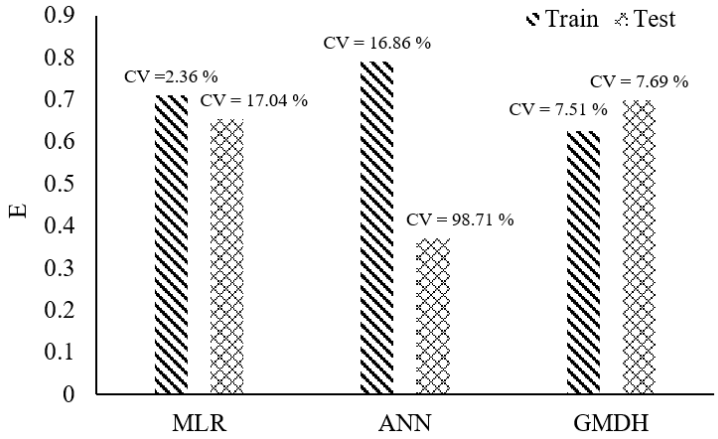

Figure 6. A summary of Nash-Sutcliffe (E) criterion along with its variation over ten replicates for three applied methods

equal accuracy and reliability levels. We believe that the lower number of predictors in developed PTFs can decrease the cost (in terms of the money, time, and labor work) for their further applications.

Regarding the included soil properties as predictors in PTF from GMDH procedure, it seems that the information from soil textural (*cc*), compactness (D_b), and porosity (θ_{fs}) as well as soil structure (θ_d) had been taken part to characterize K_s. The first three are well-documented by several researchers up to now (Brakensiek et al., 1984; Campbell and Shiozawa, 1992; Cosby et al., 1984; Dane and Puckett, 1994; Puckett et al., 1985; Saxton et al., 1986; Vereecken et al., 1990; Wösten, 1997; Wösten et al., 1999). The last one, soil moisture deficit from its optimum value for sampling (θ_d), is the one that we are introducing as an alternative and proxy indicator for soil structure. The idea came from the fact that soil sampling for water movement characterization should be conducted without or at least with the minimum changes in soil structure since it is a key factor in pore size distribution and water movement. In this regard, soil sampling at field capacity or near field capacity is strongly advised (Grossman et al., 2002) to prevent damage to soil structure. However as stated before, in most cases especially in arid or semiarid regions, the desirable soil water content for soil sampling is not present and soil is usually sampled beyond its optimum water content. Therefore, we further introduce θ_d as an alternative and proxy indicator of soil structural. In this regard, we believe that θ_d selection by GMDH, as a powerful indicators identification tool (Pachepsky and Rawls, 1999) in soil functions modeling, supports our idea in introducing θ_d as an indicator for soil structure.

3.3. GMDH *vs*. well-known exponential form of K_s-related PTFs

At the last step, we also calibrated Eq. 1 ($K_s = a \times e^b$) based on the dataset presented within this study using three different sets of input variables. The first set included all available variables, while the second set included the most applied input variables from literature (Brakensiek et al., 1984; Campbell and Shiozawa, 1992; Cosby et al., 1984; Dane and Puckett, 1994; Puckett et al., 1985; Saxton et al., 1986; Vereecken et al., 1990; Wösten, 1997; Wösten et al., 1999) containing D_b, *cc*, *si*, *sa*, and *OC*. The third set contained the most effective input variables selected by GMDH algorithm including *cc*, θ_d, D_b and D_p, and θ_{fs}. Equation 1 was selected because all the previous studies are based on the exponential form to estimate K_s.

Analogous to other PTFs, we first calibrated Eq. 1 using the train data and then applied the PTF to the validation dataset. We also repeated the random division of full dataset into a train and a validation dataset ten times, where the calibration and validation of Eq. 1 was subsequently conducted for each replicate. The results (Table 7) revealed that applying the mostly used input variables (*cc*, *si*, *sa*, D_b, and *OC*) resulted in the lowest accuracy (*E* lower than zero) for both the train and the validation datasets. While applying either all available input variables or the most effective input variables detected by GMDH algorithm (*cc*, θ_d, D_b and D_p, and θ_{fs}) showed higher accuracies for both the train dataset (*E* higher than 0.7) and the validation dataset (*E* higher than 0.6). Table 7 also shows that the GMDH algorithm effectively determined the best input variables for K_s prediction. Because using these identified variables resulted in similar or higher accuracy to that of using all available input variables, with *E* 0.73 *vs*. 0.69 for the train dataset and *E* of 0.65 vs. 0.61 for the validation dataset.

On the other hand, comparing the results for Eq. 1 and the PTFs developed by the GMDH approach revealed that the PTF developed with the GMDH showed better accuracy ($E = 0.70$) for the validation dataset compared to Eq. 1 using all available input variables ($E = 0.61$) or using the most effective input variables identified by GMDH ($E = 0.65$).

Table 7. The efficiency analysis of Eq. 1 using three different input variables.

Subset	Input variable type	n	E	RMSE	Reliability (%)
Train	All variables	94	0.694	0.121	86.84
	Mostly applied	94	-0.224	0.226	51.33
	GMDH identified	94	0.730	0.115	96.79
Test	All variables	40	0.610	0.135	83.87
	Mostly applied	40	-0.504	0.245	64.24
	GMDH identified	40	0.653	0.127	89.77

4. Conclusion

The study intended to provide a performance analysis of three different methods including multiple regression (MR), artificial neural network (ANN), and group method of data handling (GMDH) to produce several pedo-transfer functions (PTFs) to predict soil saturated hydraulic conductivity (K_s). The following conclusions are drawn from our work:

- Accuracy: Based on the Nash-Sutcliffe criterion we conclude that the GMDH resulted in more accurate predictions of K_s than MR and ANN.
- Reliability: Based on the coefficient of variation we conclude that GMDH also resulted in more reliable predictions of K_s than applying the MR or the ANN approaches.
- The comparison between MR and ANN showed that MR resulted in more accurate and more reliable K_s predictions than ANN.
- GMDH efficiently reduced the number of input variables since this subset of variables resulted in the same accuracy detected for the calibrated PTF based on an exponential form using all input variables.

One important aspect to be mentioned here is that although the ANN and ML models' architectures were developed with 10 input variables, the GMDH algorithm worked also on full input set. Contrary to ANN and ML, GMDH makes internal variable selection and then automatically includes in models only the predictors, which significantly affects model quality. Thus, all models started by working on the same input space; however, GMDH had a natural dimension reduction within its modelling process. This decreases the natural property of GMDH and is not caused by different data supplied to the models. The lower predictive performance by the ANN model with full set of predictors also demonstrates the superiority of GMDH in this study.

From practical implications point of view, the ability of GMDH to provide high level of accuracy and reliability in a few numbers of inputs makes it interesting for operational use for soil hydraulic characterization. The GMDH-based PTFs are, therefore, more cost-effective and time-saving options for the users, because information on clay content, bulk density, θ_d, and θ_{fs} is always available or can be obtained at relatively low cost. The decrease in number of predictor indices can be translated directly into less time-consuming sampling and analysis, something that is useful for large soil surveys or monitoring programs. However, to acknowledge the transferability and limitations, it should be considered that the dataset utilized for this study includes soil from northwestern Iran with four textural classes (sandy loam, sandy clay loam, loam and clay loam,) and a variety of physical conditions characteristic of semi-arid agricultural landscapes. The apparent ability of the developed PTFs to produce good estimates suggests robustness in this respect for GMDH, but caution should be exercised relative to their transferability elsewhere (on sand, clays, or organic soils not considered by the dataset and /or under different climatic regimes and management histories). Additional testing and perhaps recalibration would be required before application of these PTFs to other soil environments outside the range we have for our dataset.

Overall, this work shows that GMDH provides a potent and parsimonious modeling strategy of K_s prediction, with performance efficiency as well as methodological robustness across the empirical approaches and more intricate machine-learning architectures.

Despite its good performance in this study, GMDH also has some limitations that need to be mentioned. The statistical approach may be sensitive to data size and not sufficiently capture complex nonlinear interactions given the sample sizes. Its self-organizing structure selection might also produce another model structure if used at a different region or combined with larger soil datasets, although it has the advantage of variable parsimony. The conclusions about the GMDH performance must be taken in relation to this particular sampling size and the soil conditions here considered.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

References

Agyare, W.A., Park, S., Vlek, P., 2007. Artificial neural

network estimation of saturated hydraulic conductivity. Vadose Zone Journal 6(2), 423-431.

Aimrun, W., Amin, M., 2009. Pedo-transfer function for saturated hydraulic conductivity of lowland paddy soils. Paddy and Water Environment 7(3), 217-225.

Albalasmeh, A., Mohawesh, O., Gharaibeh, M., Deb, S., Slaughter, L., & El Hanandeh, A. (2022). Artificial neural network optimization to predict saturated hydraulic conductivity in arid and semi-arid regions. Catena, 217, 106459.

Alvarez-Acosta, C., Lascano, R.J., Stroosnijder, L., 2012. Test of the Rosetta pedotransfer function for saturated hydraulic conductivity. Open Journal of Soil Science 2(03), 203.

Arshad, R., Sayad, G., Mazlum, M., Jafarnejadi, A., Mohammadi Safarzadeh, V., 2010. Pedo-transfer functions application to estimate the infiltration rate of the soil using neural network and linear regression methods. Journal of Crop Improvement 2(5), 55-62.

Arshad, R.R., Sayyad, G., Mosaddeghi, M., Gharabaghi, B., 2013. Predicting saturated hydraulic conductivity by artificial intelligence and regression models. ISRN Soil Science 2013.

Bouma, J., 1989. Using soil survey data for quantitative land evaluation, Advances in soil science. Springer, pp. 177-213.

Brakensiek, D., Rawls, W., Stephenson, G., 1984. Modifying SCS hydrologic soil groups and curve numbers for rangeland soils. American Society of Agricultural Engineers.

Brooks, R.H., Corey, A.T., 1966. Properties of porous media affecting fluid flow. Journal of the Irrigation and Drainage Division 92(2), 61-90.

Campbell, G., Shiozawa, S., 1992. Prediction of hydraulic properties of soils using particle-size distribution and bulk density data. Indirect methods for estimating the hydraulic properties of unsaturated soils. University of California, Riverside, 317-328.

Campbell, G.S., 1974. A simple method for determining unsaturated conductivity from moisture retention data. Soil science 117(6), 311-314.

Christiaens, K., Feyen, J., 2001. Analysis of uncertainties associated with different methods to determine soil hydraulic properties and their propagation in the distributed hydrological MIKE SHE model. Journal of Hydrology 246(1), 63-81.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water resources research 20(6), 682-690.

Dane, J., Puckett, W., 1994. Field soil hydraulic properties based on physical and mineralogical information, Proceedings of the international workshop on indirect methods for estimating the hydraulic properties of unsaturated soils. University of California, Riverside, pp. 389-403.

Doussan, C., Ruy, S., 2009. Prediction of unsaturated soil hydraulic conductivity with electrical conductivity. Water Resources Research 45(10).

Elbisy, M. S. (2025). Predictive Modeling of Saturated Hydraulic Conductivity using Machine Learning Techniques. Engineering, Technology & Applied Science Research, 15(2), 21348-21355.

Farlow, S.J., 1984. Self-organizing methods in modeling: GMDH type algorithms, 54. CrC Press.

Flint, A.L., Flint, L.E., 2002. 2.2 Particle Density. Methods of Soil Analysis: Part 4 Physical Methods (methodsofsoilan4), 229-240.

Gee, G.W., Or, D., 2002. 2.4 Particle-size analysis. Methods of soil analysis. Part 4, 255-293.

Ghanbarian-Alavijeh, B., Liaghat, A., Sohrabi, S., 2010. Estimating saturated hydraulic conductivity from soil physical properties using neural networks model. World Acad. Sci. Eng. Technol 4, 108-113.

Grossman, R., Reinsch, T., 2002. 2.1 Bulk density and linear extensibility. Methods of Soil Analysis: Part 4 Physical Methods (methodsofsoilan4), 201-228.

Grossman, R., Reinsch, T., Dane, J., Topp, G., 2002. Methods of soil analysis. Part 4. Physical methods. Methods of soil analysis: Parth 4. Physical methods.

Gupta, R., Rudra, R., Dickinson, W., Patni, N., Wall, G., 1993. Comparison of saturated hydraulic conductivity measured by various field methods. Transactions of the ASAE 36(1), 51-55.

Hecht-Nielsen, R., 1990. Solution for a distributed hydrological model and applications. Neurocomputing, Addison-Wesley, Reading, MA, 89-93.

Herbst, M., Diekkrüger, B., Vanderborght, J., 2006. Numerical experiments on the sensitivity of runoff generation to the spatial variation of soil hydraulic properties. Journal of Hydrology 326(1), 43-58.

Islam, N., Wallender, W.W., Mitchell, J.P., Wicks, S., Howitt, R.E., 2006. Performance evaluation of methods for the estimation of soil hydraulic parameters and their suitability in a hydrologic model. Geoderma 134(1), 135-151.

Jabro, J., 1992. Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data. Transactions of the ASAE 35(2), 557-560.

Julia, M.F., Monreal, T.E., del Corral Jiménez, A.S., Meléndez, E.G.a., 2004. Constructing a saturated hydraulic conductivity map of Spain using pedotransfer functions and spatial prediction. Geoderma 123(3), 257-277.

Klute, A., Dirksen, C., 1986. Hydraulic conductivity and diffusivity: Laboratory methods. Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods (methodsofsoilan1), 687-734.

Kosugi, K.i., 1996. Lognormal distribution model for unsaturated soil hydraulic properties. Water Resources Research 32(9), 2697-2703.

Logsdon, S., Berli, M., Horn, R., 2013. Quantifying and

modeling soil structure dynamics. Soil Science Society of America.

Luk, K., Ball, J.E., Sharma, A., 2000. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. Journal of Hydrology 227(1), 56-65.

Mallants, D., Jacques, D., Tseng, P.-H., van Genuchten, M.T., Feyen, J., 1997a. Comparison of three hydraulic property measurement methods. Journal of hydrology 199(3-4), 295-318.

Mallants, D., Mohanty, B.P., Vervoort, A., Feyen, J., 1997b. Spatial analysis of saturated hydraulic conductivity in a soil with macropores. Soil Technology 10(2), 115-131.

Masís-Meléndez, F., Deepagoda, T.C., de Jonge, L.W., Tuller, M., Moldrup, P., 2014. Gas diffusion-derived tortuosity governs saturated hydraulic conductivity in sandy soils. Journal of Hydrology 512, 388-396.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma 109(1), 41-73.

Merdun, H., Çınar, Ö., Meral, R., Apan, M., 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil and Tillage Research 90(1), 108-116.

Minasny, B., Hopmans, J., Harter, T., Eching, S., Tuli, A., Denton, M., 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. Soil Science Society of America Journal 68(2), 417-429.

Mohanty, B., Kanwar, R.S., Everts, C., 1994. Comparison of saturated hydraulic conductivity measurement methods for a glacial-till soil. Soil Science Society of America Journal 58(3), 672-677.

Møller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural networks 6(4), 525-533.

Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., Vereecken, H., 2017. A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. Earth Syst. Sci. Data Discuss. 2017, 1-25.

Moosavi, A. A., Nematollahi, M. A., & Omidifard, M. (2024). Comparing machine learning approaches for estimating soil saturated hydraulic conductivity. PloS one, 19(11), e0310622.

Mozaffari, H., Moosavi, A. A., & Nematollahi, M. A. (2024). Predicting saturated and near-saturated hydraulic conductivity using artificial neural networks and multiple linear regression in calcareous soils. Plos one, 19(1), e0296933.

Mozaffari, H., Pakjoo, M., Nematollahi, M. A., Forouzan, S., & Moosavi, A. A. (2025). Predicting Soil Hydraulic Conductivity: A Review of Artificial Neural Networks Applications. Artificial Intelligence Applications for a Sustainable Environment, 441-462.

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water resources research 12(3), 513-522.

Naderianfar, M. (2025). Developing a simple artificial intelligence fuzzy-based model for estimating saturated hydraulic conductivity of soil. Scientific Reports, 15(1), 28476.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology 10(3), 282-290.

Nelson, D., Sommers, L.E., 1982. Total carbon, organic carbon, and organic matter. Methods of soil analysis. Part 2. Chemical and microbiological properties (methodsofsoilan2), 539-579.

Neyshabouri, M.R., Rahmati, M., Doussan, C., Behroozinezhad, B., 2013. Simplified estimation of unsaturated soil hydraulic conductivity using bulk electrical conductivity and particle size distribution. Soil research 51(1), 23-33.

Neyshaboury, M.R., Rahmati, M., Alavi, S.A.R., Rezaee, H., Nazemi, A., 2015. Prediction of unsaturated soil hydraulic conductivity using air permeability: Regression approach. Indian Journal Of Agricultural Research 49(6).

Nimmo, J.R., Perkins, K.S., 2002. 2.6 Aggregate Stability and Size Distribution. Methods of soil analysis: Part 4, 317-328.

Pachepsky, Y., Rawls, W., Gimenez, D., Watt, J., 1998. Use of soil penetration resistance and group method of data handling to improve soil water retention estimates. Soil and Tillage Research 49(1), 117-126.

Pachepsky, Y.A., Rawls, W., 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Science Society of America Journal 63(6), 1748-1757.

Pachepsky, Y.A., Timlin, D., Varallyay, G., 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Science Society of America Journal 60(3), 727-733.

Paige, G.B., Hillel, D., 1993. Comparison of three methods for assessing soil hydraulic properties. Soil Science 155(3), 175-189.

Parasuraman, K., Elshorbagy, A., Si, B.C., 2006. Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. Soil Science Society of America Journal 70(6), 1851-1859.

Puckett, W., Dane, J., Hajek, B., 1985. Physical and mineralogical data to determine soil hydraulic properties. Soil Science Society of America Journal 49(4), 831-836.

Rahmati, M., Oskouei, M. M., Neyshabouri, M. R., Walker, J. P., Fakherifard, A., Ahmadi, A., & Mousavi, S. B. (2015). Soil moisture derivation using triangle method in the lighvan watershed, north western Iran. Journal of soil science and plant nutrition, 15(1), 167-178.

Rahmati, M., 2017. Reliable and accurate point-based prediction of cumulative infiltration using soil readily available characteristics: a comparison between GMDH, ANN, and MLR. Journal of Hydrology On Press.

Rahmati, M., Neyshaboury, M.R., 2016. Soil Air Permeability Modeling and Its Use for Predicting Unsaturated Soil Hydraulic Conductivity. Soil Science Society of America Journal 80(6), 1507-1513.

Rahmati, M., Neyshabouri, M. R., Mohammadi-Oskooei, M., Fakheri-Fard, A., & Ahmadi, A. (2020). Characterizing soil infiltration parameters using field/laboratory measured and remotely-sensed data. Environmental Resources Research, 8(2), 129-146.

Reynolds, W., Elrick, D., 1985. In situ measurement of field-saturated hydraulic conductivity, sorptivity, and the α-parameter using the guelph permeameter. Soil Science 140(4), 292-302.

Reynolds, W., Elrick, D., Youngs, E., Amoozegar, A., Booltink, H., Bouma, J., 2002. 3.4 Saturated and field-saturated water flow parameters. Methods of soil analysis, Part 4, 797-801.

Or, D., Keller, T., & Schlesinger, W. H. (2021). Natural and managed soil structure: On the fragile scaffolding for soil functioning. Soil and Tillage Research, 208, 104912.

Sarmadian, F., Taghizadeh-Mehrjardi, R., 2014. Estimation of infiltration rate and deep percolation water using feed-forward neural networks in Gorgan Province. Eurasian Journal of Soil Science 3(1), 1.

Saxton, K., Rawls, W.J., Romberger, J., Papendick, R., 1986. Estimating generalized soil-water characteristics from texture. Soil Science Society of America Journal 50(4), 1031-1036.

Schaap, M.G., Leij, F.J., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. Soil and Tillage Research 47(1), 37-42.

Schaap, M.G., Leij, F.J., Van Genuchten, M.T., 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Science Society of America Journal 62(4), 847-855.

Sharghi, F., Bauke, S. L., Rahmati, M., Burger, D. J., Vereecken, H., & Amelung, W. (2025). Soil infiltration variability across diverse soil reference groups, textures, and landuse types. Geoderma, 463, 117550.

Sirkin, R.M., 2006. Two-sample t test. In: R.M. Sirkin (Ed.), Statistics for the Social Sciences Thousand Oaks, Calif.: Sage Publications. xxi, London, New Delhi, pp. 271-358.

Spychalski, M., Kaźmierowski, C., Kaczmarek, Z., 2007. Estimation of saturated hydraulic conductivity on the basis of drainage porosity. Electronic Journal of Polish Agricultural Universities 10(1), 04.

Suleiman, A., Ritchie, J., 2001. Estimating saturated hydraulic conductivity from soil porosity. Transactions of the ASAE 44(2), 235.

Tietje, O., Hennings, V., 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69(1-2), 71-84.

van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil science society of America journal 44(5), 892-898.

Vereecken, H., Maes, J., Feyen, J., 1990. Estimating unsaturated hydraulic conductivity from easily measured soil properties. Soil Science 149(1), 1-12.

Vereecken, H., Schnepf, A., Hopmans, J., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M., Amelung, W., Aitkenhead, M., 2016. Modeling soil processes: Review, key challenges, and new perspectives. Vadose zone journal 15(5).

Vereecken, H., Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., ... & Zhang, Y. (2022). Soil hydrology in the Earth system. Nature Reviews Earth & Environment, 3(9), 573-587.

Webster, M., 2006. Merriam-Webster online dictionary.

Weihermüller, L., Lehmann, P., Herbst, M., Rahmati, M., Verhoef, A., Or, D., ... & Vereecken, H. (2021). Choice of pedotransfer functions matters when simulating soil water balance fluxes. Journal of Advances in Modeling Earth Systems, 13(3), e2020MS002404.

Wösten, J., 1997. Pedotransfer functions to evaluate soil quality. Developments in Soil Science 25, 221-245.

Wösten, J., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90(3), 169-185.

Yamaç, S. S., Negiş, H., Şeker, C., Memon, A. M., Kurtuluş, B., Todorovic, M., & Alomair, G. (2022). Saturated hydraulic conductivity estimation using artificial intelligence techniques: a case study for calcareous alluvial soils in a semi-arid region. Water, 14(23), 3875.

Zhao, C., Shao, M.a., Jia, X., Nasir, M., Zhang, C., 2016. Using pedotransfer functions to estimate soil hydraulic conductivity in the Loess Plateau of China. Catena 143, 1-6.