Estimating saturated hydraulic conductivity using different well-known pedotransfer functions

A. Haghverdi[1], H. S. Özturk[2], S. Ghodsi[3] and T. Tunçay[4]

[1] Dept. of Agriculture, Payame Noor University (PNU), Neishabour, Iran, amirhaghverdii@gmail.com

[2] Dept. of Soil Science and Plant Nutrition, Faculty of Agriculture, Ankara University, hozturk@agri.ankara.edu.tr

[3] Dept. of Geology, Faculty of science, Birjand University, Birjand, Iran, somaie.ghodsi@gmail.com

[4] Earth Science Application and Research Center, Ankara University, Gölbaşı, Ankara, tuncay@agri.ankara.edu.tr

ABSTRACT

Saturated hydraulic conductivity (K_s) is a key input factor in agricultural modeling tools. The direct methods of measuring this property, either in situ or in the lab, are time consuming, expensive and impractical in large scales. Pedotransfer functions (PTFs), which relate easy collected data to hydraulic properties, can use to solve these problems. In this research, the performance of some already published well-known PTFs including Jabro, Puckett., Rosseta and NeuroTheta were compared to in terms of the accuracy of the estimation of log K_s. Beside these models a local PTF, Turkey-PTF, was derived to analyze the effect of the using local data for PTF accuracy. For this purpose, 91 undisturbed soil samples with 2 replications were collected from the surface layer (0-30cm). The K_s of samples were measured in the laboratory using by laboratory permeameter instrument. The soil texture, dry bulk density and organic matter were measured with common methods and used as input parameters. There was a trend to underestimate the K_s in all well-known PTFs. Regarding to the root mean squared (RMSE) and correlation coefficient (*r*) values, the better result among the well-known PTFs belonged to NeuroTheta PTF. In general, in the competition of the neural network, NN, based and regression based PTFs, the lowest error and the better performance belonged to NN based ones. However, the supremacy of local PTF, Turkey-PTF, was remarkable comparing with well-known PTFs. Regarding to the moderate performances of the other PTFs, it would be suggested to use already published well-known PTFs for new soils and new areas cautiously.

INTRODUCTION

Soils are naturally heterogeneous, and this heterogeneity affects their hydraulic properties even in a small scale (Parasuraman et al., 2006). Therefore, one of the most critical research topics in the soil science is how to estimate the hydraulic properties in a fast and integrated way regardless to the scale of the research area. Saturated hydraulic conductivity, K_s, is one of the most important properties of soil, which should be identified as an input parameter in description and mathematical simulation of the most of the models related to the water and soil engineering (Ghanbarian-Alavijeh et al., 2010).

Although in theory, it would be best to measure soil hydraulic parameters in the lab or field to identify the variability in space and over time sufficiently, from the practical point of view, direct determination of hydraulic properties is still costly and time consuming (Aimrun and Amin, 2009; Nemes et al., 2005). As the result over the past 2 decades, more than the effort which scientists have been done to develop of new experimental techniques to determine soil hydraulic properties, they focused to establish pedotransfer function, PTF. The term PTF, which firstly introduced at the latest of 80s, includes mathematical relationships for estimating soil hydraulic properties from more readily available data, e.g., texture, soil organic matter, OM, and bulk density, BD.

While most of the PTFs have been developed to estimate water retention curve, there are a few amount of researches available in literature which tried to derive the PTF for predicting K_s (Vereecken et al., 2010). Like most of the other research area, PTFs firstly derived by regression based method, but after a while data mining based PTFs replace the old ones in a great extent due to the fact that in almost all cases they work better than the regression based ones (Ghanbarian-Alavijeh et al., 2010). Aimrun and Amin (2009) derived PTF to estimate K_s of lowland paddy soils. Ghanbarian-Alavijeh et al. (2010) estimated the K_s from soil physical properties using neural network model, NN. In another study, Agyare et al. (2007) used NN to estimate K_s. It was observed by them that good data distribution, range, and amounts are prerequisites for good NN estimation and, therefore, data preprocessing is important for NN.

Further revaluation of already published PTFs plays an important role for identifying the reliability of them otherwise the PTFs become useless after derivation. For this reason, this study was carried out to address the reliability of some famous well-known PTFs for estimating K_s comparing with a local data based PTF.

MATERIAL AND METHODS

SAMPLING AND LABOR EXPERIMENTS

A dataset, containing 91 samples, was gathered from the different parts of Turkey. The soil texture triangle and the physical properties of soil samples are shown in Figure 1 and Table 1, respectively.

The disturbed and undisturbed (100 cm^3) soil samples were collected from the surface soil (0-30 cm). Two undisturbed samples were taken from each location by using a dedicated soil sample ring kit (Eijkelkamp Agrisearch Equipment, Giesbeek, The Netherlands). The soil samples were collected after the first rainfalls following the long dry period. The moisture content of samples was lower to their field capacity at the time of sampling. To reduce the effect of management practices on soil structure and in turn K_s, the locations of sampling were selected from long term fallow lands. Undisturbed samples were used to measure BD and hydraulic properties, but the disturbed samples were used to identify the rest of properties, including particle-size distribution and OM.

Particle-size distribution was determined by the hydrometer method (Gee and Bauder, 1986). OM content was determined by the modified Walkley and Black method (USDA, 1982). The K_s of samples were measured with 3 repetitions in the laboratory using by laboratory permeameter instrument. Once the hydraulic characteristics were determined the soil cores were used to determine dry BD (Blake and Hartge, 1986).

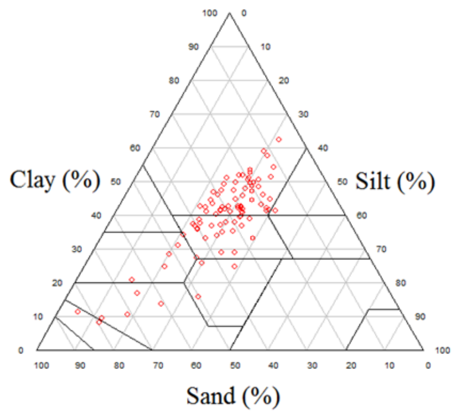

Figure 1.Soil texture of the dataset used in this study. clay corresponds to 0-2 µm, silt to 2-50 µm and sand to 50-2000 µm.

Table 1: Physical characteristics of soil samples

	Max	Min	Average	SD
Sand (%)	83.6	5.9	31.9	15.1
Silt (%)	41.4	5.2	28.6	6.5
Clay (%)	62.2	8.0	39.5	11.1
OM (%)	3.85	0.01	1.16	0.66
BD (g cm^{-3})	1.63	0.93	1.19	0.15

Table 2. Characteristics of the well-known PTFs used in this study

PTF	Input variables	Software	Source
Jabro	sand, silt, clay, BD	SOILPAR2	Jabro (1992)
Puckett	sand, silt, clay	SOILPAR2	Puckett et al. (1985)
Rosetta	sand, silt, clay, BD	Rosetta	Schaap et al. (2001)
NeuroTheta	sand, silt, clay	NeuroTheta	Minasny and McBratney (2003)
Turkey-PTFs	sand, silt, clay, BD, OM		

PTF

Four well known PTFs, Jabro, Puckett, Rosetta and NeuroTheta, were used to estimate K_s in this study. Table 2 shows the PTFs, input variables and corresponding softwares. The input variables (sand, silt and clay contents, SSC, and BD) are varied among the models.

Furthermore, a local regression based PTF, Turkey-PTF, was derived using 70% of samples as developments subset and 30% as test subset. Beside the similar input predictors that used for the other PTFs, soil organic matter, OM, was used as an extra predictor in Turkey-PTF.

PERFORMANCE CRITERIA

In this study, the root mean square of error (RMSE), correlation coefficient (r) and mean bias error (MBE) were selected to evaluate the performances of the PTFs. Since logarithmic values of K_s are considered, the corresponding RMSE and MBE statistics are dimensionless.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(E_i - M_i)^2}{n}} \quad (1)$$

$$r = \frac{\sum_{i=1}^{n}(M_i - \overline{M})(E_i - \overline{E})}{\sqrt{\sum_{i=1}^{n}(M_i - \overline{M})^2 \sum_{i=1}^{n}(E_i - \overline{E})^2}} \quad (2)$$

$$\mathrm{MBE} = \sum_{i=1}^{n}\frac{E_i - M_i}{n} \quad (3)$$

where M_i is the measured value, E_i is the estimated value for ith sample, $\overline{M}$ is the average of measured values, $\overline{E}$ is the average of estimated values and n is the number of samples. The program IRENE (www.isci.it/tools), a data analysis tool designed to provide easy access to model evaluation techniques, was used for identifying the statistics.

RESULT AND DISCUSSION

The scatter plots of measured versus estimated log K_s for Jabro, Puckett, Rosetta, NeuroTheta and Turkey-PTFs, and the performance statistics of them are illustrated in Figure 2 and Table

3, respectively. In general, the performances of well-known PTFs were not good. In all four well-known PTFs, most of the points lied under 1:1 line, which means the PTFs underestimate the K_s for most of the samples. The fact of the trend to underestimate can also be proved by observing negative MBE values of all PTFs in Table 3. The absolute values of MBE varied between 0.32-2.63. Estimated log K_s by Jabro and Rosetta had a negative correlation with measured log K_s while the Puckett and NeuroTheta showed a positive correlation. The absolute values of correlation varied between 0.13-0.50, which are somehow similar to the study of Agyare et al (2007) who used the NN model for estimating K_s using PTF concept, and reported the range of 0.29-0.41 for correlation of measured with estimated K_s.

Among well-known PTFs, the lowest RMSE value belongs to Rosetta, 1.61, and the highest belongs to Puckett, 2.8. However, regarding to the both RMSE and *r* values, NeuroTheta shows the better performance among the used PTFs. Regarding to the fact that the lowest RMSE belonged to Rosetta and the better performance to NeuroTheta, a conclusion could be drawn; the performance of ANN based PTFs, NeuroTheta and Rosetta, was slightly better than the regression based PTFs, Jabro and Puckett.

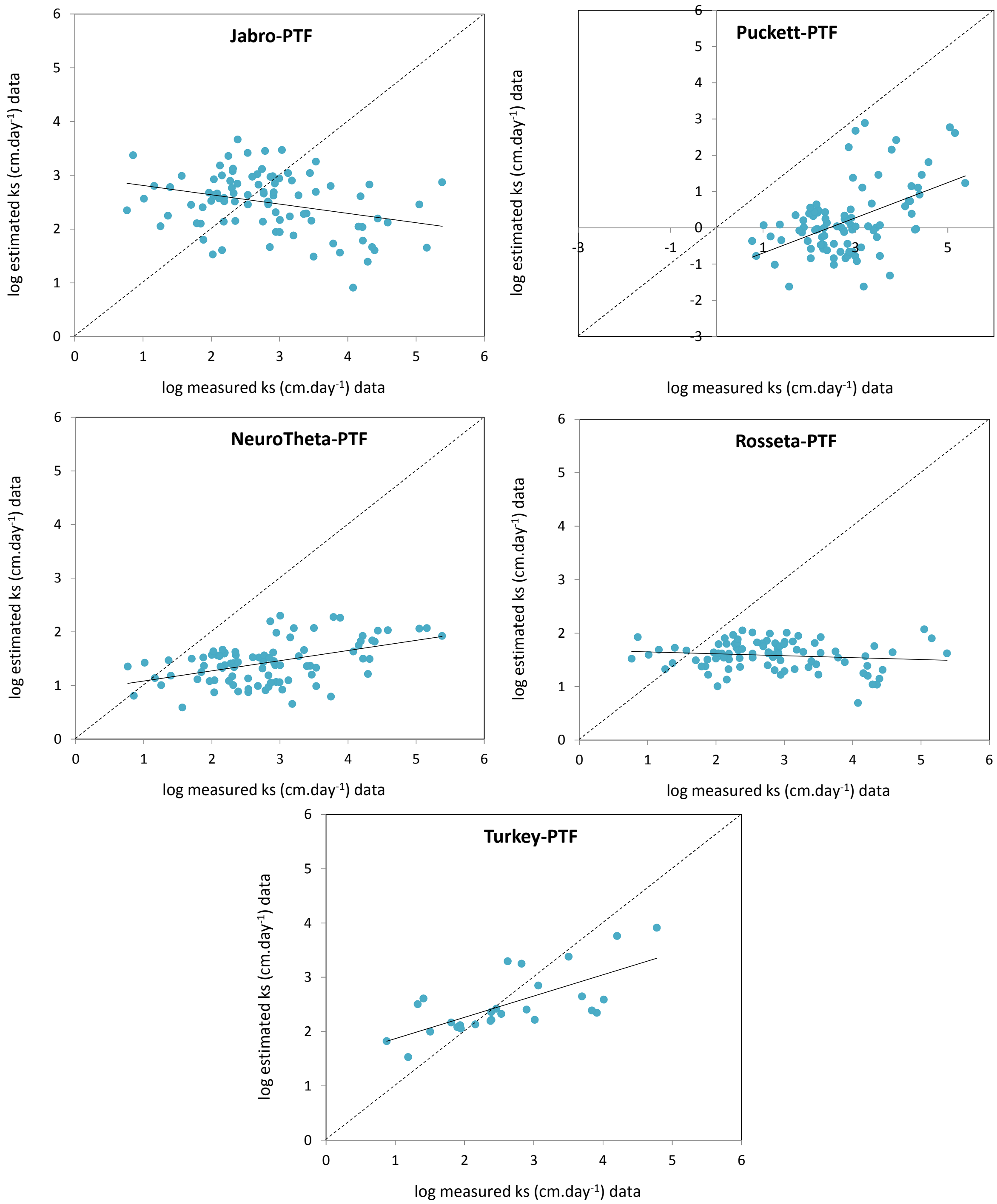

Figure 2. Scatter plots of measured versus estimated log k_s (mm.day^{-1}) for all PTFs

While all of the well-known PTFs worked poor to moderate, the performance of Turkey-PTF was good. The RMSE of Turkey PTF is almost twice times lower than that of the best well-known PTF, and the correlation coefficient of Turkey-PTF is almost 1.5 times greater than that of the best well known-PTF. Figure 2 also shows the remarkable difference between the performance of well-known PTFs and local PTF. Li et al. (2007) showed that The PTFs obtained from local dataset have superior performances in predicting the soil hydraulic parameters, compared to PTFs derived from large datasets, which in turn confirm the limitation of applying PTFs developed from one region to other regions. Importance of using the local data for deriving K_s PTF also has been emphasized by the result of the study of Parasuraman et al. (2006). They used Rosetta and field scale NN model to estimate K_s using basic soil data for two different datasets, and the RMSE values of test data were 0.48-0.87 and 0.19-0.26. The highest values of RMSE for both datasets in the study of Parasuraman et al. (2006) belonged to Rosetta in comparison to the local field scale derived PTFs. It can be concluded that for modeling of K_s by PTF concept, using local data (if they are available) is the best option. Aimrun and Amin (2009) reported the correlation 0.68 between measured and estimated K_s after transformation, which is close to the performance of Turkey-PTF in our study. They used local data for derivation, as well.

In addition, using the OM as an extra input predictor in the study of the Aimrun and Amin (2009) and the present one could be effective in higher correlation of measured and estimated K_s comparing with the well-known PTFs in the current study which only utilized sand, silt, clay and BD as the predictors.

Table 3: Evaluation statistics of PTFs

	RMSE	MBE	*r*
Jabro	1.29	-0.32	-0.31
Puckett	2.80	-2.63	0.50
Neurotheta	1.63	-1.39	0.48
Rosetta	1.61	-1.23	-0.13
Turkey	0.74	-0.11	0.69

REFERENCES

Agyare, W.A., Park, W.A., and Vlek. P.L.G. (2007) ‘Artificial neural network estimation of saturated hydraulic conductivity’, *Vadose Zone Journal*, 6, 423–431.

Aimrun, W., and Amin M.S.M. (2009) ‘Pedo-transfer function for saturated hydraulic conductivity of lowland paddy soils’. *Paddy Water Environmental,* 7:217–225.

Blake, G.R., and Hartge. K.H. (1986) Bulk density. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1, second ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI, pp. 363-375.

Gee, G.W., and Bauder. J.W. (1986) Particle size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1, second ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI, pp. 383-411.

Ghanbarian-Alavijeh, B., Liaghat, A.M., and Sohrabi S. (2010) ‘Estimating saturated hydraulic conductivity from soil physical properties using neural networks model’, *World Academy of Science, Engineering and Technology*, 62, 131-136.

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

Li, Y., Chen, D., White, R.E., Zhu, A., and Zhang. J., (2007) ‘Estimating soil hydraulic properties of Fengqiu County soils in the North China Plain using pedo-transfer functions’, *Geoderma* 138, 261–271.

Minasny, B., and McBratney, A.B. (2003) ‘NeuroTheta, pedotransfer functions for predicting soil hydraulic properties for Australian soil’, Australian Centre for Precision Agriculture, The University of Sydney. http://www.usyd.edu.au/su/agric/acpa.

Nemes, A., Rawls, W.J., and Pachepsky, Y.A. (2005) ‘Influence of organic matter on the estimation of saturated hydraulic conductivity’, *Soil Science Society of America Journal*, 69, 1330–1337.

Parasuraman, K., Elshorbagy, A., and Cheng Si, B. (2006), ‘Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles’, *Soil Science Society of America Journal*, 70, 1851–1859.

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

Schaap, M.G., Leij, F.J., and Van Genuchten, M.Th. (2001) ‘Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions’, *Journal of Hydrology*, 251, 163–176.

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., and van Genuchten, M.Th. (2010) 'Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A Review', *Vadose Zone Journal*, 9: 1-26.

USDA, (1982) Procedures for collecting soil samples and methods of analysis for soil survey. Soil Survey Investigations. Report No.1.