Studding of Number of Dataset on Precision of Estimated Saturated Hydraulic Conductivity

M. Siosemarde and M. Byzedi

Abstract—Saturated hydraulic conductivity of Soil is an important property in processes involving water and solute flow in soils. Saturated hydraulic conductivity of soil is difficult to measure and can be highly variable, requiring a large number of replicate samples. In this study, 60 sets of soil samples were collected at Saqhez region of Kurdistan province-IRAN. The statistics such as Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Bias Error (MBE) and Mean Absolute Error (MAE) were used to evaluate the multiple linear regression models varied with number of dataset. In this study the multiple linear regression models were evaluated when only percentage of sand, silt, and clay content were used as inputs. The R, RMSE, MBE and MAE values of the 50 dataset for method (SSC) were calculated 0.925, 15.29, -1.03 and 12.51, and for method (SSC+Bd), were calculated 0.927, 15.28, -1.11 and 12.92, respectively, for relationship obtained from multiple linear regressions on data. Also the R, RMSE, MBE and MAE values of the 10 dataset for method (SSC) were calculated 0.725, 19.62, -9.87 and 18.91 and for method (SSC+Bd), were calculated 0.618, 24.69, -17.37 and 22.16, respectively, which shows when number of dataset increase, precision of estimated saturated hydraulic conductivity, increases.

Keywords—dataset, precision, saturated hydraulic conductivity, soil and statistics.

I. INTRODUCTION

The saturated hydraulic conductivity of soil is a measurement parameter of the soil's ability to transmit water when submitted to a hydraulic gradient that is defined by Darcy's law. The hydraulic conductivity depends on the soil grain size, the structure of the soil matrix, the type of soil fluid, and the relative amount of soil saturation present in the soil matrix. The important properties relevant to the solid matrix of the soil include pore size distribution, pore shape, tortuosity, specific surface, and porosity [11 & 14]. Numerous field and laboratory techniques have been developed to directly measure this quality [1 & 25].

Since, direct measurement of hydraulic conductivity is time consuming and costly, the cost-effectiveness of obtaining soil hydraulic conductivity can be improved by using indirect methods, which allow the estimation of hydraulic conductivity from more easily measured parameters [2, 3, 5, 13, 17, 20, 24, 26, 27, 28, 29, 32, 33, 34, 35 & 36].

Knowledge about the variability of soil properties is probably as old as the soil classification system [12 & 30]. Variability in the hydraulic properties of soil units has been studied by several researchers [7, 12 & 15]. Field observations show that the hydraulic properties of soils vary significantly with spatial location even within a given soil type [12 & 22].

Hazen (1982) proposed the relationship between saturated hydraulic conductivity and soil particle diameter such that 10% of all soil particles are finer (smaller) by weight [6 & 9]. Shepherd (1989) extended Hazen’s research by performing power regression analysis [21]. Puckett et al. sampled six soils at seven different locations in the Alabama lower coastal plain [18], and used regression analysis to determine that percentage of clay sized particles was the best predictor of Ks. Rawls and Brakensiek used field data across the U.S. to develop a regression equation that relates porosity, and the percentages of sand and clay-sized particles in the sample to Ks [19]. Jabro estimated Ks from grain-size and bulk density data [10]. Sperry and Peirce (1995) developed a linear model to estimate Ks based on grain size, shape, and porosity, also they performed an evaluation by comparing the measured hydraulic conductivity values of different porous materials with those determined from their own equations as well as the equations of Hazen, Kozeny-Carman and Alyamani and Sen [23]. Cronican and Gribb (2004) developed a multiple linear regression for southeastern U.S. sandy soils based on regional soil data [6]. Han et al. (2008) developed a new model to estimate saturated hydraulic conductivity from soil structural properties derived from water retention curve [8]. Jadczyzs and Nied¥wiecki (2005) reported that the lower content of both silt and organic matter and lower values of bulk density had increased Ks [11]. The results showed that the hydraulic conductivities calculated by the USBR and Slitcher methods are in all cases lower than for the other methods [4, 16 & 31].

Although many researches carried out for prediction of saturated hydraulic conductivity from soil texture and particle diameter but in this research an attempt has been made to study the effect of dataset number on precision of estimated saturated hydraulic conductivity.

II. MATERIALS AND METHODS

In this study, 60 sets of soil samples were collected to measure hydraulic conductivity based on parameters of soil texture at location of Saqhez region of Kurdistan province-IRAN. Standard methods were applied to investigate parameters of soil texture (clay, silt and sand). Soil texture...
was classified according to the International Society of Soil Science (ISSS) classification system. The textural classification of soils was clay, silty clay, silty clay loam, silty loam and loam. Bulk density was determined from the mass of dry soil contained in 120 cm² steel sampling cylinders. The samples were oven-dried at 105°C for 48 hours. The results were calculated as g/cm³. The content of clay, silt & sand and values of bulk density and saturated hydraulic conductivity are summarized in Table 1. The mean contents of clay, silt and sand were 36.00, 50.45 and 13.12 [g (100g)⁻¹], respectively, also the mean values of bulk density and saturated hydraulic conductivity were 1.31 (g/cm³) and 50.40 (cm/day).

In this study saturated hydraulic conductivity was measured by the falling-head method. The falling-head method is very similar to the constant head methods in its initial setup; however, the advantage to the falling-head method is that can be used for both fine-grained and coarse-grained soils. The falling-head method is very similar to the constant head methods in its initial setup; without maintaining a constant pressure head and saturated hydraulic conductivity was measured from the bottom so that air could escape from the upper surface. The water is then allowed to flow through the soil without maintaining a constant pressure head and saturated hydraulic conductivity was measured when the rate of head loss was constant.

In this study multiple linear regression models obtained by different number of dataset including 50, 40, 30, 20 and 10 series of dataset and the obtained models were evaluated using the other 10 series of dataset. The multiple linear regression models were developed when only percentage of sand, silt, and clay content (SSC) were used as inputs, and when SSC and bulk density, Bd, (SSC+Bd) were used as inputs. The equation 5, 6, 7, 8 and 9 for \( K_s \), were obtained from multiple linear regressions on 50, 40, 30, 20 and 10 series of dataset when (SSC) were used as inputs respectively, and the equation 10, 11, 12, 13 and 14 for \( K_s \) were obtained from multiple linear regressions on 50, 40, 30, 20 and 10 series of dataset when (SSC+Bd) were used as inputs data respectively.

\[
K_s = \frac{\sum_{i=1}^{n} (O_i - \bar{O}) (P_i - \bar{P})}{\sum_{i=1}^{n} (O_i - \bar{O})^2 (P_i - \bar{P})^2}
\]

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}
\]

\[
MBE = \frac{\sum_{i=1}^{n} (P_i - O_i)}{n}
\]

\[
MAE = \frac{\sum_{i=1}^{n} |P_i - O_i|}{n}
\]

The following equations for \( K_s \), saturated hydraulic conductivity (cm/day), were obtained from multiple linear regressions on different dataset when only percentage of sand, silt, and clay content (SSC) were used as inputs, and when SSC and bulk density, Bd, (SSC+Bd) were used as inputs. The equation 5, 6, 7, 8 and 9 for \( K_s \), were obtained from multiple linear regressions on 50, 40, 30, 20 and 10 series of dataset when (SSC) were used as inputs data respectively, and the equation 10, 11, 12, 13 and 14 for \( K_s \) were obtained from multiple linear regressions on 50, 40, 30, 20 and 10 series of dataset when (SSC+Bd) were used as inputs data respectively.

\[
K_s = -82.38 - 14.19(Bd) + 0.19(C) + 1.84(Si) + 4.96(S)
\]

\[
K_s = -184.43 - 24.35(Bd) + 0.89(C) + 3.06(Si) + 6.22(S)
\]

\[
K_s = -290.58 - 33.33(Bd) + 2.08(C) + 4.21(Si) + 7.42(S)
\]

\[
K_s = -265.50 - 30.33(Bd) + 1.94(C) + 3.77(Si) + 6.96(S)
\]

\[
K_s = -83.57 - 163.42(Bd) + 1.37(C) + 3.84(Si) + 6.95(S)
\]

\[
K_s = -115.48 + 0.01(C) + 1.95(Si) + 5.11(S)
\]

\[
K_s = -237.07 + 1.19(C) + 3.20(Si) + 6.4(S)
\]

\[
K_s = -410.10 + 2.97(C) + 4.90(Si) + 8.09(S)
\]

\[
K_s = -386.56 + 2.88(C) + 4.53(Si) + 7.65(S)
\]

\[
K_s = -534.13 + 4.60(C) + 5.90(Si) + 8.69(S)
\]

Where C is the percentage of clay-sized particles, Si is the percentage of silt-sized particles, S is the percentage of sand-sized particles and Bd is the soil bulk density (g/cm³).

For this case-study, the statistics of different multiple linear regression models related to various number of dataset, when only three (SSC) inputs were used and when four (SSC+Bd) inputs were used in predicting \( K_s \), are presented in Table 2 and Table 3 respectively. The best model of multiple linear regressions was selected based on the four statistics e.g. Correlation Coefficient (R), Root Mean Square Error (RMSE), Mean Bias Error (MBE) and Mean Absolute Error (MAE).

Values of statistics presented in Table 2 and Table 3 show when SSC and bulk density, (SSC+Bd) were used as inputs.
The research investigated to study the effect of dataset number on precision of estimated saturated hydraulic conductivity. The results presented in Table 2 and Table 3 showed that when number of dataset increase, the statistics of RMSE, MBE and MAE decrease and statistic of R increases, therefore the results of this study showed the equations were obtained from multiple linear regressions on more dataset estimated \( K_s \), better (less error) than equations from less dataset.

### IV. Conclusion

The research investigated to study the effect of dataset number on precision of estimated saturated hydraulic conductivity. The results of multiple linear regressions showed when SSC and bulk density, \( (SSC+BD) \) were used as inputs data estimated \( K_s \), almost such when only percentage of sand, silt, and clay content \( (SSC) \) were used as inputs data. Values of statistics presented in Tables 2 and 3 resulted in under prediction (negative MBE) for both methods of \( SSC \) and \( SSC+BD \).

The statistics of \( R \), RMSE, MBE and MAE decrease and statistic of \( R \) increases, therefore the results of this study showed the equations were obtained from multiple linear regressions on more dataset estimated \( K_s \), better (less error) than equations from less dataset.

### ACKNOWLEDGMENT

The authors wish to thank the Sama technical and vocational training school, Islamic Azad University, Mahabad Branch, for supporting this research.

### REFERENCES


