J. Indian Water Resour. Soc.,
Vol. 37, No. 4, Oct., 2017

MODELING OF UNSATURATED HYDRAULIC CONDUCTIVITY OF SOIL

Parveen Sihag[1], N.K. Tiwari[2] and Subodh Ranjan[3]

ABSTRACT

This paper investigates the potential of machine learning approaches, i.e. Multi-linear Regression (MLR), Support Vector Machine (SVM) and Gaussian Process (GP) regression of the unsaturated hydraulic conductivity of soil. Dataset consisting of 46 samples was used and obtained from the laboratory experiments. Out of 46 samples arbitrarily selected 32 samples were used for training, whereas remaining 14 were used for testing the models. Input data set consisted of sand, rice husk ash, fly ash, suction head, bulk density and moisture content whereas unsaturated hydraulic conductivity was considered as output. Two kernels function i.e. Pearson VII and radial based kernel functions were used with both SVM and GP regression. A comparison of results suggested that SVM approach works well than the GP and MLR approaches and it could successfully be used in the modeling of unsaturated hydraulic conductivity of soil. Sensitivity analyses suggested that percentage of sand and moisture content were important parameters in the prediction of unsaturated hydraulic conductivity of soil.

Keywords: *Unsaturated hydraulic conductivity, Multi-linear Regression, Support vector regression, Gaussian process regression.*

INTRODUCTION

In the state of irrigation and agriculture, water management is required to improve the efficiency and sustainability of agricultural systems when water is scarce (Siltecho et al. 2014). One of the tasks of water management is to compute and control water. Consequentially, these processes are dependent on the soil hydraulic conductivity which needs to be determined in the field or in the laboratory. On the other hand, direct measurement of hydraulic conductivity of soil is difficult, tedious, relatively costly, labour intensive and time-consuming (Lakzian et al.2010, Emami et al.2012, Kaikhajeh et al.2012 and Fereshte 2014). Hydraulic conductivity is one of the most important parameters for determining infiltration rates and other hydrological process (Gulser and Candemir, 2008). Thus, indirect methods using predictive approaches have been developed for estimation of hydraulic properties of a soil from easily measurable soil properties (Schuh et al. 1986). However, predictive approaches of the soil hydraulic conductivity have gained considerable attention and efforts have been made by researchers to improve the power of predictability (Skaggs et al. 2001). The predictive approaches of the soil hydraulic conductivity may be developed by different methods like multi-linear regression and artificial neural network. These methods have been recommended in the field of hydraulic conductivity as encourage results of different research papers (Tamari et al. 1996, Nosrati et al. 2012) were reported.

The relationship between unsaturated hydraulic conductivity (K_u *in cm/hr)*of soil with sand (S in %), rice husk ash (R_{ha} in %), fly ash (F_a in %), suction head (S_h in cm) , bulk density (B_d *in gm/cc)*and moisture content (M_C *in %)* is non-linear; therefore, a following functional relationship may be initially assumed:

$$K_u = p_i S^a R_{ha}{}^b F_a{}^c S_h{}^d B_d{}^e M_C{}^f \quad (1)$$

In which, p_i is the proportionality constant.

Taking logs of both sides of the equation (1) so as to reduce the equation to a linear form,

$$\log K_u = \log p_i + a \log S + b \log R_{ha} + c \log F_a + d \log S_h + e \log B_d + f \log M_C \quad (2)$$

which is a multi-linear equation with seven explanatory variables. Now to develop a multi-linear model, log K_u is taken as the output parameter and the seven explanatory variables, namely logS , log R_{ha}, logF_a, logS_h, logB_d and logM_C are taken as input parameters. The output of the multi-linear regression provided the values of p_i, a, b, c, d, e & f and, in turn, the developed equation of the form (1). The developed multi-linear regression equation is as under:

$$K_u = 0.1397 S^{0.559} R_{ha}{}^{1.1964} F_a{}^{-0.2791} S_h{}^{-0.2070} B_d{}^{5.5445} M_C{}^{-0.7306} \quad (3)$$

Within the last few decades, soft computing approaches like artificial neural network, Gaussian process, Support vector machines and M5 model tree have been used in civil and hydraulics applications (Nawari et al., 1999; Dibike et al., 2001; Cigizoglu, 2005; Kurup and Griffin, 2006; Pal, 2006; Pal and Deswal, 2008;Pal and Singh, 2014; Elbisy, 2015; Singh et al., 2016) and found to be very effective.

Keeping in the view the potential of Gaussian process and support vector machines based regression approaches; the present study explores the capabilities of these techniques in modeling of unsaturated hydraulic conductivity of soil and compares their performance with an empirical relationship developed by MLR.

Support vector Machines

This method was introduced by Vapnik (1995) and derived from statistical learning theory. The main principle of SVM is optimal separation of classes, from the separable classes SVM selects the one which have least generalisation error from an infinite number of linear classifier or set upper limit of the error which is obtained from structural risk minimisation. Thus the maximum margin between two classes could be obtained from the selected hyper plane and sum of the distances of the hyper plane from the closest point of two classes will set the maximum margin between two classes.

Vapnik (1995) anticipated ε-Support Vector Regression (SVM) by introducing another ε- insensitive loss function and it permits the concept of margin to be used for regression problems. The principle of the SVM is to discover a function

Ph.D Research Scholar[1], Associate Professor[2], Professor[3]
Department of Civil Engineering, N.I.T. Kurukshetra, Haryana

Manuscript No.: 1456

having at most ε deviation from the actual target vectors for all specified training data and it should be as flat as possible (Smola (1996)). For a given training data with k number of samples be represented by $\{\mathbf{x_i}, y_i\}$, i = 1, …, k, where $\boldsymbol{x}_i$ is input vector and y_i is the target value, a linear decision function can be represented by

$$f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b \tag{4}$$

Where $\boldsymbol{w} \in \boldsymbol{R}^N$ and b $\in$ R. $\langle \mathbf{w}, \mathbf{x} \rangle$ represents the dot product in space $\mathbf{R}^N$. In Equation 5, vector **w** determine the orientation of a discriminating plane, whereas scalar b determine the offset of the discriminating plane from the origin. A smaller value of **w** indicates the flatness of Equation (4), which can be achieved by minimising the Euclidean norm defined by $\|\mathbf{w}\|^2$ (Vapnik (1995)). Thus, an optimisation problem for regression can be written as (Smola, A. J., 1996):

$$\text{Minimise } \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{subject to } \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \le \varepsilon \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \le \varepsilon \end{cases} \tag{5}$$

The optimisation problem in Equation (5) is based on the assumption that there exists a function that provides an error on all training pairs which is less than ε. In real life problems, there may be a situation like the one defined for classification by Cortes, C. and Vapnik (1995). So, to allow some more error, slack variables ξ, ξ^* can be introduced and the optimisation problem defined in Equation (5) can be written as below to deal with infeasible constraints of the optimization problem (Smola (1996)):

$$\text{Minimise } \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{k}(\xi_i + \xi_i^*)$$

$$\text{Subject to } y_i - \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - b \le \varepsilon + \xi_i$$

$$\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b - y_i \le \varepsilon + \xi_i^* \quad \text{and} \quad \xi_i, \xi_i^* \ge 0 \text{ for all i = 1, 2,……, k.} \tag{6}$$

The optimisation problem in Equation (6) can be solved by replacing the inequalities with a simpler form determined by transforming the problem to a dual space representation using Lagrangian multipliers (Luenberger (1984)).

The Lagrangian of Equation (6) can be formed by introducing positive Lagrange multipliers $\lambda_i, \lambda_i^*, \eta_i, \eta_i^*$ i = 1,….,k and multiplying the constraint equations by these multipliers, and finally subtracting the results from the objective function. The Lagrangian for Equation (6) can now be written as:

$$L = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{k}(\xi_i + \xi_i^*) - \sum_{i=1}^{k}\lambda_i(\varepsilon + \xi_i - y_i + \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) - \sum_{i=1}^{k}\lambda_i^*(\varepsilon + \xi_i^* + y_i - \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle - b) - \sum_{i=1}^{k}(\eta_i \xi_i + \eta_i^* \xi_i^*) \tag{7}$$

The dual variables in Equation (7) have to satisfy λ_i, λ_i^*, η_i, $\eta_i^* \ge 0$. The solution of the optimisation problem involved in the design of SVM can be obtained by locating the saddle point of the Lagrange function defined in the equation (7)

The prediction problem in the equation (4) can now be written as:

$$f(\boldsymbol{x}) = \sum_{i=1}^{k}(\lambda_i^* - \lambda_i)\langle \boldsymbol{x}_i, \boldsymbol{x} \rangle + b \tag{8}$$

$$f(\boldsymbol{x}) = \sum_{i=1}^{k}(\lambda_i^* - \lambda_i) K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{9}$$

In this optimisation problem, the kernel function is computed so as to decrease the computational cost of dealing with the high dimension feature space. For a detail study of SVM readers are referred to (Vapnik (1995).

GP Regression

Rasmussen (2006) assumption on which GP regression model works are that adjoining observation should express information about each other, it is a method of specifying a prior directly over function space. The mean and covariance of Gaussian distribution are vector and matrix where as Gaussian process is over function. GP regression models are able to recognize the predictive distribution analogous to test input.

A GP is a collection of random variables, any finite number of which has a joint multivariate Gaussian distribution. Let X Y represent the domains of inputs and outputs, respectively, out of which *n* pairs (x_i, y_i) are drawn independently and identically distributed. For regression, let $y \subseteq \Re$; then, a GP on χ is defined by a mean function $\mu: \chi \to \Re$ and a covariance function $k: \chi \times \chi \to \Re$.

The main assumption of GP regression is that *y* is given by $y = f(x) + \xi$, where $\xi \sim N(0, \sigma^2)$. The symbol ~ in statistics means *sampling for*. In GP regression, for every input *a* there is an associated random variable *f (x)*, which is the value of the stochastic function *f* at that location. In this work, it is assumed that the observational error ξ is normal independent and identically distributed, with a mean value of zero ($\mu(x)=0$), a variance of σ^2 and $f(x)$ drawn from the Gaussian process on χ specified by *k*. That is,

$$Y = (y_1, ..., y_n) \sim N(0, K + \sigma^2 I),$$

where $K_{ij} = k(x_i, x_j)$, and **I** is the identity matrix.

Because $Y/X \sim \mathrm{N}\left(0, K+\sigma^2 \mathbf{I}\right)$ being normal, so is the conditional distribution of test labels specified the training and test data of $p\left(Y_* / Y, X, X_*\right)$. Then, one has $Y_* / Y, X, X_* \sim \mathrm{N}(\mu, \Sigma)$, where

$$\mu = K\left(X_*, X\right)\left(K\left(X, X\right)+\sigma^2 \mathbf{I}\right)^{-1} Y, \quad (10)$$

$$\Sigma = K\left(X_*, X_*\right)-\sigma^2 \mathbf{I}-K\left(X_*, X\right)\left(K\left(X, X\right)+\sigma^2 \mathbf{I}\right)^{-1} K\left(X, X_*\right), \quad (11)$$

If there are *n* training data and n_* test data, then $K\left(X, X_*\right)$ represents the $n \times n_*$ matrix of covariance's evaluated at all pairs of training and test data sets, and this is similarly true for the other values of $K\left(X, X\right)$, $K\left(X_*, X\right)$ and $K\left(X_*, X_*\right)$; here *X* and *Y* are the vector of the training data and training data labels y_i, whereas X_* is the vector of the test data.

A specified covariance function is required to produce a positive semi-definite covariance matrix *K*, where $K_{ij} = k\left(\mathbf{x}_i, \mathbf{x}_j\right)$. The term of the kernel function used in SVM is equivalent to the covariance function used in GP regression. With the known values of kernel *k* and degree of noise σ^2, Equations (10) and (11) would be sufficient for inference.

During the training process of GP regression models, one needs to choose an appropriate covariance function as well as its parameters. In the case of GP regression with a permanent value of Gaussian noise, a GP model can be trained by applying Bayesian inference, i.e., maximizing the marginal likelihood. This leads to the minimization of the negative log-posterior:

$$p\left(\sigma^2, k\right) = \frac{1}{2} \mathrm{y}^T \left(K+\sigma^2 \mathbf{I}\right)^{-1} \mathrm{y} + \frac{1}{2} \log\left|\mathrm{K}+\sigma^2 \mathbf{I}\right| - \log\, p\left(\sigma^2\right) - \log\, p(k) \quad (12)$$

To find the hyper parameters, the partial derivative of Equation 12 can be obtained with respect to σ^2 and *k*, and minimization can be achieved by gradient descent. For further information about GP regression and different covariance functions readers are referred to Kuss (2006).

METHODOLOGY AND DATA SET

Soil and Water Samples Characteristics

Experiments were conducted at hydraulic laboratory located in National Institute of Technology, Kurukshetra, India. The soil selected for experimentation was sand. Rice husk ash and fly ash mixed in sand at different proportions. All the measurement was conducted on predetermined initial conditions like moisture content, dry density. The moisture content of the samples was measured with the help of electric oven for 24 hr at 100°C. The samples were carefully compacted in proctor having a volume of 1000 cm^3.Properties of the material used is presented in Table 1.

Table 1: Properties of the material used for experimentation

Properties	Sand	Rice Husk Ash	Fly Ash
Specific gravity	2.48	1.89	2.07
D_{50} (mm)	0.438	0.190	0.180
C_u	3.1290	3.200	2.7333
Color	White	Black	Gray

Measurement of unsaturated soil hydraulic conductivity in the Laboratory

The unsaturated soil hydraulic conductivity (Ku) was measured in the laboratory (Figure 1) using a mini disk infiltrometer (Decagon Devices Inc.) with assistance of some calculations which are provided as excel worksheet soil (Decagon Devices Inc., 2014). The mini disk infiltrometer consists of two chambers (water reservoir and bubble chamber), which are connected via a Mariotte tube to provide a constant water pressure head of -0.5 to -7 cm (equivalent to -0.05 to -0.7 kPa). The bottom of the mini disk infiltrometer contains a porous sintered steel disk having 4.5 cm diameter and 3 mm thick. The water filled tube is placed upon the soil surface (Figure 1) resulting in water infiltrating into the soil, with the volume of water and speed of infiltration depending on the sorptivity and hydraulic conductivity of the soil. A suction head of -1 cm to -6 cm (equivalent to -0.1 to 0.6 kPa) was chosen in this study. During the measurement, the volume of the water in the reservoir chamber was recorded at regular intervals.

Calculating of unsaturated soil hydraulic conductivity (Ku)

The method proposed by Zhang (1997) is quite simple and works well for measurements of infiltration into unsaturated soil from the recorded data by mini disk infiltrometer (Decagon Devices, 2014). The method requires measuring cumulative infiltration vs. time and fitting the results with the function:

$$I = C_1 t + C_2 \sqrt{t} \quad (13)$$

Where I is the cumulative infiltration (cm), t is the time (*sec*), and C_1 (cm/sec) and C_2 (cm $sec^{-0.5}$) are parameters. C_1 is related to hydraulic conductivity and C_2 is related to soil sorptivity. The hydraulic conductivity (K) of the soil is then computed from:

$$k = \frac{C_1}{A} \quad (14)$$

Where C_1 is the slope of the curve of the cumulative infiltration vs. the square root of time (Figure 2).and A is a value relating the Van Genuchten (1980) parameters for a given soil type to the suction rate and radius of the infiltrometer disk. A is computed from:

$$A = \frac{11.65(n^{0.1}-1)\exp\left[2.92(n-1.9)\alpha h_0\right]}{(\alpha r_0)^{0.91}} \quad for \quad n \geq 1.9$$

$$= \frac{A}{11.65(n^{0.1} - 1)\exp\,[7.5(n - 1.9)\alpha h_0}{(\alpha r_0)^{0.91}} \quad for\ n \leq 1.9$$

Where n and α are the Van Genuchten parameters for the soil, r_0 is the disk radius and h_o is the suction at the disk surface. The mini disk infiltrometer infiltrates water at a suction of -0.5 to -6 cm and has a radius of 2.25 cm. The Van Genuchten parameters for the 12 texture classes were obtained from Carsel and Parrish (1988).

The experiments were conducted on mixture of Sand, rice husk ash and fly ash for different moisture content, bulk density and suction head. The characteristics of experimental data are specified in Table 2 and correlation matrix of training data is shown in Table 3

Details of GP and SVM

The GP and SVM-based regression approaches design involves the idea of kernel functions. A number of kernels are discussed in the literature, but studies suggest a better performance by radial basis kernels for different civil engineering problems (Pal and Mather, 2003; Gill et al., 2006;

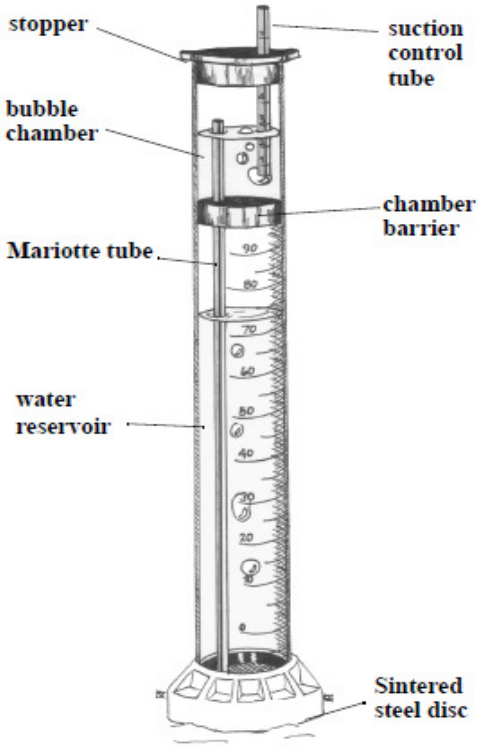

Fig. 1: Mini disc Infiltrometer (Infiltrometer User's Manual, 2014)

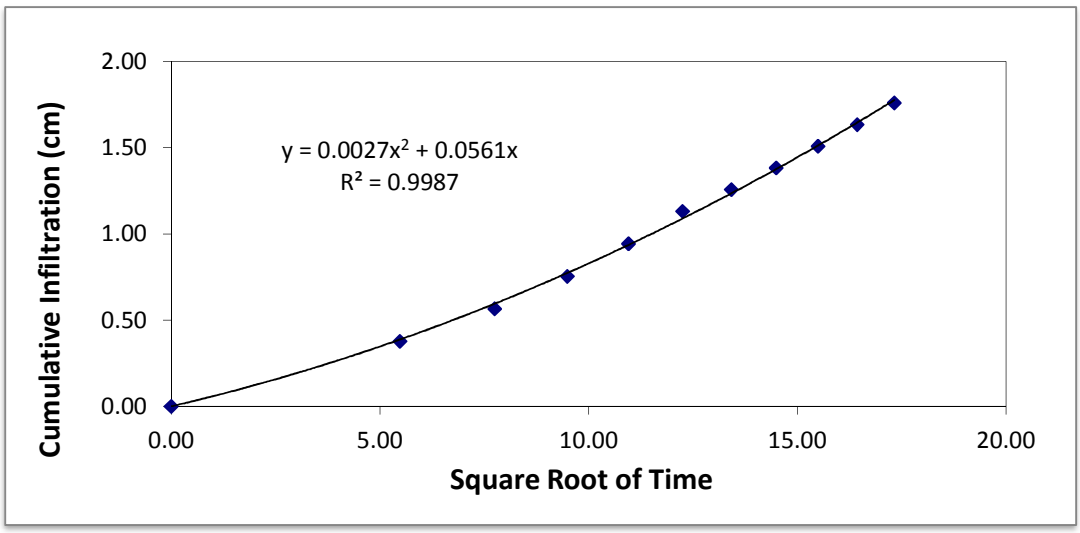

Fig. 2: Cumulative infiltration versus square root of time for a soil.

Table 2: Characteristics of train and test data used

Input parameter	Train data				Test data			
	Min	**Max**	**Mean**	**St. dev.**	**Min**	**Max**	**Mean**	**St. dev.**
S	50	90	61.25	14.973	50	90	60	14.142
R_{ha}	5	45	20.938	13.704	5	45	23.571	13.506
F_a	5	45	17.813	12.309	5	40	16.425	10.271
S_h	0.5	6	1.234	1.205	0.5	4	1.25	0.849
B_d	0.836	1.729	1.182	0.24	0.836	1.559	1.144	0.225
M_c	2	20	10.094	5.177	2	20	10.5	4.256

Table 3: Correlation Matrix of training data set

Variables	S	F_a	R_{ha}	S_h	B_d	M_c
S	**1.0000**	-0.5098	-0.6347	-0.1509	0.8150	0.0858
F_a	-0.5098	**1.0000**	-0.3412	-0.3776	0.0366	-0.0498
R_{ha}	-0.6347	-0.3412	**1.0000**	0.5040	-0.9233	-0.0490
S_h	-0.1509	-0.3776	0.5040	**1.0000**	-0.3805	-0.0036
B_d	0.8150	0.0366	-0.9233	-0.3805	**1.0000**	0.2611
M_c	0.0858	-0.0498	-0.0490	-0.0036	0.2611	**1.0000**

Pal and Singh, 2010). Recently, Üstün et al., (2006) proposed a universal Pearson VII function-based kernel and suggested it to be an alternative to the linear, polynomial and radial basis function kernels. The present study uses the radial basis kernel ($e^{-\gamma\left|\mathbf{x_i}-\mathbf{x_j}\right|^2}$) and the Pearson VII function kernel $\left(1\Big/\left[1+\left(2\sqrt{\left\|\mathrm{x}_i-\mathrm{x}_j\right\|^2}\sqrt{2^{(1/\omega)}-1}\Big/\sigma\right)^2\right]^{\omega}\right)$, where γ,σ and ω are kernel-specific parameters. The parameter σ controls the Pearson width, whereas ω is the tailing factor of the peak when the Pearson VII function is used for curve-fitting purposes.

The use of GP- and SVM-based modeling approaches requires the establishments of suitable user-defined parameters, as the accuracies of both regression approaches are mainly dependent on the choice of these parameters. In addition to the choice of a kernel and kernel-specific parameters, the SVM requires the establishment of the regularization parameter *C* and the size of the error-insensitive zone ε, whereas GP regression requires the determination of the optimal values of the level of Gaussian noise (added to the diagonal of the Covariance Matrix).

In the present study, a manual method (carrying out a large number of trials by using different combinations of user-defined parameters with both modeling approaches) was used to select user-defined parameters (i.e., *C*, γ,σ, ω, ε and Gaussian noise). For SVM, multiple trials were also carried out to find a suitable value of the error-insensitive zone with a fixed value of *C* and kernel-specific parameters. The optimal values of various user-defined parameters are selected in such a way so as to minimize the root mean square error and maximize the coefficient of correlation. The same kernel-specific parameters were used for both GP regression and SVM. Table 4 provides the optimal values of the user-defined parameters for both radial basis function and Pearson VII kernel-based GP regression and SVM. To evaluate the performance of both modeling approaches and empirical relation (MLR), coefficient of correlation and root mean square error (RMSE) values were used.

Table 4: User-defined parameters using Gaussian process and SVM based regression for RBF and PUK kernel.

	Support vector machines	**Gaussian process regression**
RBF kernel	$C = 2,\ \gamma = 2$	Gaussian noise $= 0.01,\ \gamma = 2$
PUK kernel	$C = 2,\ \omega = 2,\ \sigma = 1$	Gaussian noise $= 0.01,\ \omega = 2,\ \sigma = 1$

RESULT AND DISCUSSION

Figure 3, 4, 5 and 6 provide plots between observed and predicted unsaturated hydraulic conductivity of soil by Pearson VII and RBF kernel function based GP and SVM regression as well as multiple linear regression (MLR) approaches using training and testing data set. Results from

MLR, GP and SVM regression approaches are in accordance with the observed values. Results of the test data set from Table 5 indicate that the performance of RBF kernel function based SVM is best among both GP and MLR approaches. Correlation coefficient value of 0.9339 (RMSE = 2.9771 cm/hr) was achieved by RBF kernel function based SVM. RBF kernel function based SVM suggests a better performance in comparison to RBF kernel based GP approach. The Correlation coefficient value of 0.8934 (RMSE =3.6606cm/hr) was achieved by Pearson VII kernel based SVM approach. In comparison to both GP and SVM regression approaches, SVM approach performance was quite better (Table 5) The correlation coefficient value of 0.8020 (RMSE = 5.1767 cm/hr) was achieved by MLR approach which is not agreeable. Keeping in view the improved performance of RBF kernel function based SVM modeling approaches, a graph between test data set number and unsaturated hydraulic conductivity (K_u) is plotted (Figure 7). It can be inferred from this figure that predicted value provided by RBF kernel function based SVM were in very close proximity to the observed unsaturated hydraulic conductivity and predicted unsaturated hydraulic conductivity found to follow the same patterns of observed unsaturated hydraulic conductivity.

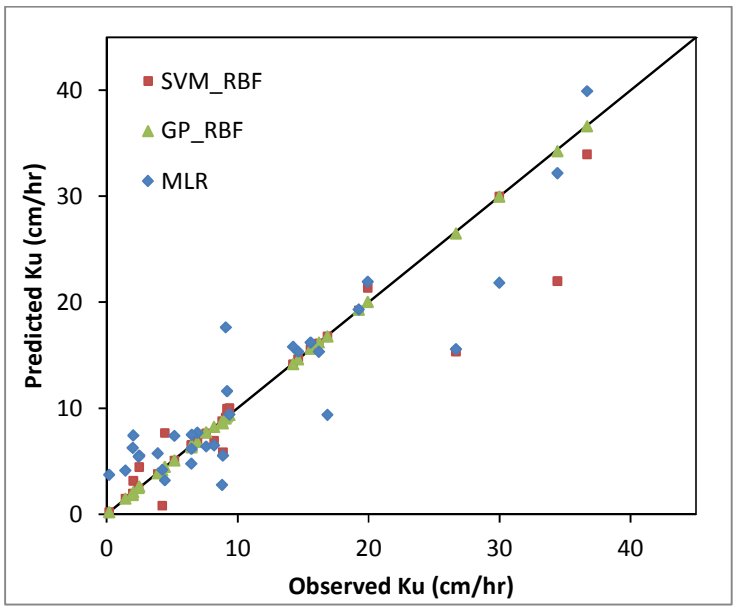

Fig. 3: Observed vs. predicted values of unsaturated hydraulic conductivity using MLR, SVM and GP regression with RBF kernel of training data

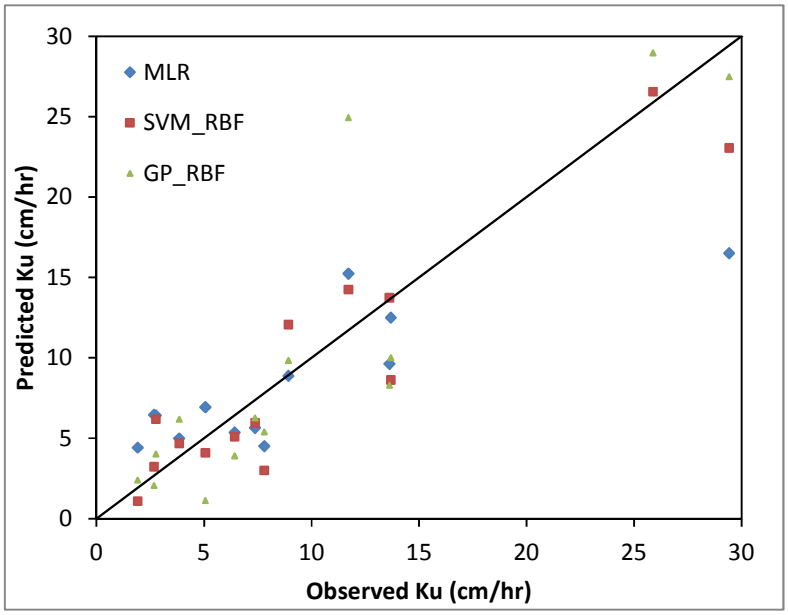

Fig. 4: Observed vs. predicted values of unsaturated hydraulic conductivity using MLR, SVM and GP regression with RBF kernel of testing data

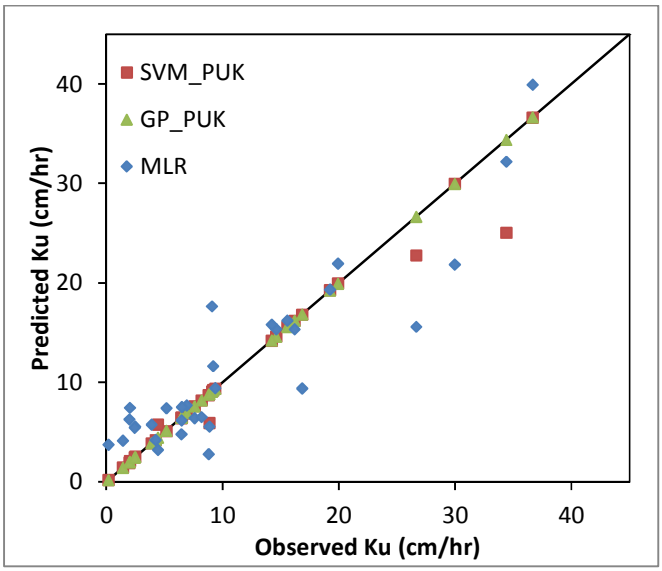

Fig. 5: Observed vs. predicted values of unsaturated hydraulic conductivity using MLR, SVM and GP regression with PUK kernel of training data

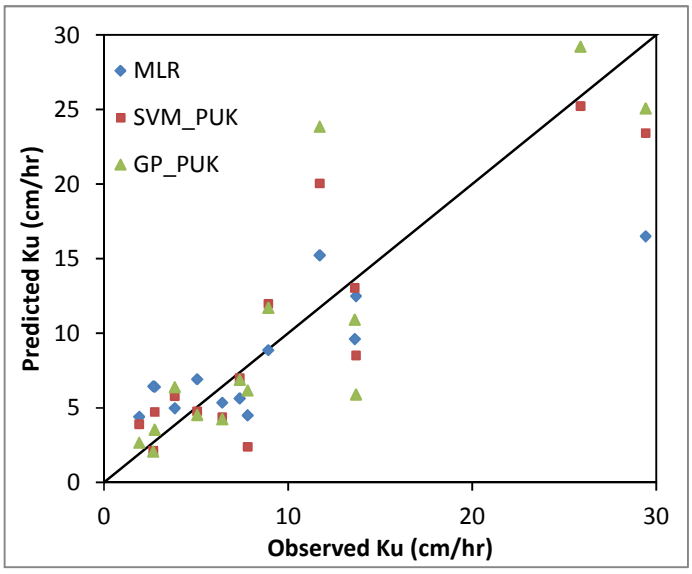

Fig. 6: Observed vs. predicted values of unsaturated hydraulic conductivity using MLR and SVM regression with PUK kernel of testing data.

Table 5: Detail of performance evaluation parameters using GP, SVM and MLR of training and testing data set

Approaches	Training data set		Testing dataset	
	Correlation coefficient	Root mean squared error (cm/hr)	Correlation coefficient	Root mean squared error (cm/hr)
Gaussian process regression with PUK kernel	1	0.0099	0.8671	4.3829
Gaussian process regression with RBF kernel	1	0.0947	0.8831	4.3708
SVM with PUK kernel	0.9835	1.8796	0.8934	3.6606
SVM with RBF kernel	0.9487	3.2085	0.9339	2.9771
MLR (empirical equation)	0.9089	4.4690	0.8020	5.1767

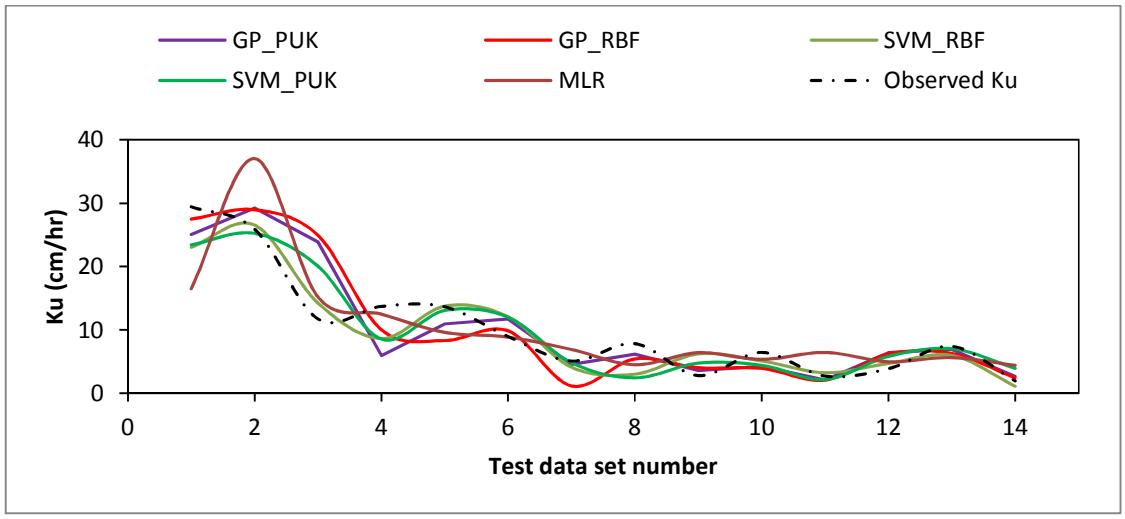

Fig.7: Variation in predicted values of K_u using the different regression approach in comparison to observed values of K_u

SENSITIVITY ANALYSIS

Sensitivity analysis was carried out to determine the most significant input parameter in unsaturated hydraulic conductivity of soil. For this, the RBF kernel based SVM regression performing also best with the data set, was used. A different set of training data was created by removing one input parameter at a time and results were reported in terms of the coefficient of correlation and root mean square error (RMSE) with the training data set. Results from Table 6 suggest that moisture content and percentage of sand of the soil have significant role in predicting the unsaturated hydraulic conductivity of soil in comparison to other input parameter.

Table 6: Sensitivity analysis using RBF kernel based SVM regression

Input combination	Input parameter removed	SVM regression	
		Coefficient of correlation	Root mean square error (cm/hr)
S, R_{ha}, F_a, S_h, B_d, M_c		0.9487	3.2085
R_{ha}, F_a, S_h, B_d, M_c	S	0.9586	2.7536
S, F_a, S_h, B_d, M_c	R_{ha}	0.9461	3.2934
S, R_{ha}, S_h, B_d, M_c	F_a	0.9477	3.2342
S, R_{ha}, F_a, B_d, M_c	S_h	0.943	3.3581
S, R_{ha}, F_a, S_h, M_c	B_d	0.9507	3.1336
S, R_{ha}, F_a, S_h, B_d	M_c	0.896	4.3505

CONCLUSION

This paper investigates the potential of MLR, SVM and GP based regression approach in predicting the unsaturated hydraulic conductivity of soil. From the comparison of performance evaluation parameters, it was found that RBF based SVM approach worked well (CC=0.9339, RMSE=2.9771 cm/hr) in comparison to GP and MLR approaches for this data set. One important conclusion was that SVM worked well with both the kernels based regression and Pearson VII based SVM (CC=0.8934, RMSE=3.6606 cm/hr) regression also gives better performance in predicting unsaturated hydraulic conductivity. The Correlation coefficient value of 0.8020 (RMSE = 5.1767 cm/hr) was achieved by MLR approach which is not agreeable. Results of sensitivity analysis concluded that moisture content and percentage of sand were most important parameter in the measurement of unsaturated hydraulic conductivity of soil.

REFERENCES

1. *Carsel, R.F. and Parrish, R.S., 1988. Developing joint probability distributions of soil water retention characteristics. Water Resources Research, 24(5), pp.755-769.*
2. *Cigizoglu, H. K., 2005. Application of generalized regression neural networks to intermittent flow forecasting and estimation. Journal of Hydrologic Engineering, 10, 336–341.*
3. *Devices, D., 2014. Mini disk infiltrometer user's manual, Version 9. Decagon Devices, Pullman, WA.*
4. *Dibike, Y. B., Velickov, S., Solomatine, D. P. and Abbott, M. B., 2001. Model induction with support vector machines: Introduction and applications, Journal of Computing in Civil Engineering, 15, 208-216.*
5. *Elbisy, M.S., 2015. Support Vector Machine and regression analysis to predict the field hydraulic conductivity of sandy soil. KSCE Journal of Civil Engineering, 19(7), pp.2307-2316.*
6. *Emami, H., SHorafa, M. and Neyshabouri, M.R., 2012. Evaluation of Hydraulic Conductivity at Inflection Point of Soil Moisture Characteristic Curve as a Matching*

Point for some Soil Unsaturated Hydraulic Conductivity Models. JWSS-Isfahan University of Technology, 16(59), pp.169-182.

7. *Fereshte, F.H., 2014. Evaluation of Artificial Neural Network and Regression PTFS in Estimating Some Soil Hydraulic Parameters. ProEnvironment/ProMediu, 7(17).*
8. *Gill, M. K., Asefa, T., Kemblowski, M. W., Makee, M., 2006. Soil moisture prediction using Support Vector Machines. Journal of the American water resources association.; Volume 42, Issue 4, 1033-1046.*
9. *Gülser, C. and Candemir, F., 2008. Prediction of saturated hydraulic conductivity using some moisture constants and soil physical properties. Proceeding Balwois.*
10. *Kalkhajesh, Y.K., Arshad, R.R., Amerikhan, H., Sami, M., 2012. Multiple linear regression, artificial neural network (MLP, RBF) and ANFIIS models for modeling the saturated hydraulic conductivity of tropical region soils (a case study: Khuzestan province:southwest Iran). International journal of Agriculture: research and Review 2(3), pp. 255-265.*
11. *Kurup, P. U. and Griffin, E. P., 2006. Prediction of soil composition from CPT data using general regression neural network. Journal of Computing in Civil Engineering, 20(4), 281–289.*
12. *Kuss, M., 2006. Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning, PhD thesis, TechnischenUniversität Darmstadt.*
13. *Lakzian, A., AVAL, M.B. and Gorbanzadeh, N., 2010. Comparison of pattern recognition, artificial neural network and pedotransfer functions for estimation of soil water parameters. NotulaeScientiaBiologicae, 2(3), p.114.*
14. *Luenberger, D.G. and Ye, Y., 1984. Linear and nonlinear programming (Volume 2). Reading, MA: Addison-wesley.*
15. *Nawari, N.O., Liang, R. and Nusairat, J., 1999. Artificial intelligence techniques for the design and analysis of deep foundations. Electronic Journal of Geotechnical Engineering, 4, pp.1-21.*
16. *Nosrati, K.F., Movahedi, N.S., Hezarjaribi, A., Roshani, G.A. and Dehghani, A.A., 2012. Using Artificial Neural Networks to estimate saturated hydraulic conductivity from easily available soil properties. Electronic journal of Soil Managment and Sustainable Production, 2 (1), pp. 95-110.*
17. *Pal, M. and Deswal, S., 2008. Modeling pile capacity using support vector machines and generalized regression neural network. Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 134(7), 1021-124.*
18. *Pal, M., 2006. Support vector machines-based modelling of seismic liquefaction potential. International Journal for Numerical and Analytical Methods in Geomechanics, 30, 983-996.*
19. *Pal, M., and Mather, P. M., 2003. Support vector classifiers for land cover classification, Map India, New Delhi, www.gisdevelopment.net/technology/rs/pdf/23.pdf.*
20. *Pal, M., Singh, N.K. and Tiwari, N.K., 2010. Support vector regression based modeling of pier scour using field data. Engineering Applications of Artificial Intelligence, Volume 24, Issue 5, pp.911-916.*
21. *Pal, M., Singh, N.K. and Tiwari, N.K., 2014. Kernel methods for pier scour modeling using field data. Journal of Hydroinformatics, 16(4), pp.784-796.*
22. *Rasmussen C. E., Williams C. K. I., 2006. Gaussian Processes for Machine Learning. The MIT Press, Cambridge, MA.*
23. *Schuh, W.M. and Bauder, J.W., 1986. Effect of soil properties on hydraulic conductivity-moisture relationships. Soil Science Society of America Journal, 50(4), pp.848-855.*
24. *Siltecho, S., Hammecker, C., Sriboonlue, V., Clermont-Dauphin, C., Trelo-Ges, V., Antonino, A.C.D. and Angulo-Jaramillo, R., 2014. Use of field and laboratory methods for estimating unsaturated hydraulic properties under different land-use. Hydrology and Earth System Sciences, 11(6), pp.6099-6137.*
25. *Singh, B.K., Tiwari, N.K. and Singh, K.K., 2016. Support vector regression based modelling of trapping efficiency of silt ejector. Journal of Indian Water Resources Society, 36 (1),pp.41-49.*
26. *Skaggs, T.H., Arya, L.M., Shouse, P.J. and Mohanty, B.P., 2001. Estimating particle-size distribution from limited soil texture data. Soil Science Society of America Journal, 65(4), pp.1038-1044.*
27. *Smola, A. J., 1996. Regression estimation with support vector learning machines. Master's Thesis, TechnischeUniversitätMünchen, Germany.*
28. *Tamari, S., Wösten, J.H.M. and Ruiz-Suarez, J.C., 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. Soil Science Society of America Journal, 60(6), pp.1732-1741.*
29. *Üstün, B., Melssen, W. J., Buydens, L. M. C., 2006. Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. Chemometrics and Intelligent Laboratory Systems, Volume 81, Issue 29 – 40.*
30. *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), pp.892-898.*
31. *Vapnik, V. N., 1995. The Nature of Statistical Learning Theory. New York: Springer-Verlag.*
32. *Zhang, R., 1997. Determination of soil sorptivity and hydraulic conductivity from the disk infiltrometer. Soil Science Society of America Journal, 61(4), pp.1024-1030.*