ORIGINAL ARTICLE

Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity

Parveen Sihag[1] · Sahar Mohsenzadeh Karimi[2] · Anastasia Angelaki[3]

Abstract

Hydraulic conductivity of soil reveals its influencing role in the studies related to management of surface and subsurface flow, e.g. irrigation and drainage projects, and solute mass transport models. Direct measurements of hydraulic conductivity have many difficulties due to spatial variation of the property in the field. Pertaining to this problem, in this study, estimation models have been developed using machine learning methods (M5 tree model and random forest model) in an attempt to estimate the accurate values of unsaturated hydraulic conductivity related to basic soil properties (clay, silt and sand content, bulk density and moisture content). Data set was collected from the experimental measurements of cumulative infiltration using mini disc infiltrometer at the study area (Kurukshetra, India). A multivariate nonlinear regression (MNLR) relationship was derived, and the performance of this model was compared with the machine learning-based models. The evaluation of the results, based on statistical criteria (R^2, RMSE, MAE), suggested that random forest regression model is superior in accurate estimations of the unsaturated hydraulic conductivity of field data relative to M5 model tree and MNLR.

Keywords Unsaturated hydraulic conductivity · M5 model tree · Random forest · Multivariate nonlinear regression

Introduction

It is essential to estimate the hydraulic features of the soil because of their considerable role in dam, hydrological cycle, irrigation system, drainage system and groundwater flow-related studies. The saturated hydraulic conductivity of soil is an important property which signifies the characterization of subsurface flow behaviour, and it largely affects the characteristics of infiltration of water through soil. Hydraulic conductivity is strongly influenced by the compacting behaviour, density and water content of the soil. The design and

feasibility of irrigation and drainage projects require accurate determination of hydraulic conductivity for the efficient water management related to surface as well as subsurface flow. Hydraulic conductivity of soil is traditionally estimated for small samples in the laboratory or by using different infiltrometers in the field. The direct measurement and accurate determination of hydraulic conductivity are difficult, tedious and time-consuming due to temporal and spatial variabilities when hydrological estimations are required for huge areas (Arshad et al. 2013). Therefore, indirect methods involving predictive estimations have received a considerable popularity and are widely adopted in order to provide reasonable predictability of hydraulic properties of soils in relation to basic measurable soil properties (Al-Sulaiman and Aboukarima 2016). Thus, for numerous hydrological model functions, soil hydraulic properties are predicted from more simply accessible proxy variables such as texture of soil, bulk density or organic carbon content (Jarvis et al. 2013). Many predictive computing methods such as multiple linear regression (MLR), artificial neural network (ANN), support vector machines (SVM) and adaptive neuro-fuzzy inference system (ANFIS) have been used to improve the estimation precision of hydraulic conductivity of soil.

✉ Parveen Sihag
  parveen12sihag@gmail.com

  Sahar Mohsenzadeh Karimi
  mohsenzadeh.sahar21@gmail.com

  Anastasia Angelaki
  anaggel@agr.uth.gr

[1] Department of Civil Engineering, National Institute of Technology Kurukshetra, Kurukshetra, Haryana, India

[2] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

[3] Department of Agriculture, Crop Production and Rural Environment, University of Thessaly, Volos, Greece

Springer

Most of the studies available in the literature discussed the application of ANN and SVM as predictive models for the estimation of hydraulic conductivity of soil (Agyare et al. 2007; Erzin et al. 2009; Rogiers et al. 2012; Das et al. 2012; Sihag 2018). Arshad et al. (2013) compared the performance of radial basis function neural networks (RBFNN), multi-layer perceptron neural networks (MLPNN), ANFIS and MLR to estimate the saturated hydraulic conductivity based on soil texture and bulk density. They reported ANFIS as a powerful estimation tool relative to ANN and MLR. Ekhmaj (2010) developed MLR and ANN models in order to predict the steady infiltration rate, and the outcomes yielded better predictions with ANN model relative to MLR. Elbisy (2015) implemented genetic algorithm in order to determine the optimum SVM parameters and investigated the performance of three kernel functions (linear, radial basis and sigmoid) in determining field hydraulic conductivity of sandy soil having easily measurable soil parameters as input variables. The study yielded RBF kernel-based SVM as a powerful method for the indirect estimations of hydraulic conductivity in comparison with other methods. Al-Sulaiman and Aboukarima (2016) successfully implemented ANN model for the accurate relationship of hydraulic conductivity with eight input soil parameters (sand, silt, clay, soil electric conductivity, sodium absorption ratio, organic matter, initial soil water content and bulk density of soil). In a study conducted on field infiltration data, Sihag et al. (2017a) suggested a novel nonlinear regression-based infiltration model developed from Kostiakov modified model for the location of NIT Kurukshetra (India) which yielded better estimations of infiltration rate than some popular conventional models. In a laboratory study conducted on synthetic soil samples by varying the percentages of soil mixture (sand, rice husk ash, fly ash), moisture content, bulk density and suction head, cumulative infiltration was estimated by using machine learning approaches (multiple nonlinear regression, support vector machines, Gaussian process regression) as well as conventional infiltration models (Sihag et al. 2017b). The study resulted in accurate predictions with Gaussian process regression (GPR) approach relative to other models. In a similar type of laboratory data, and Tiwari et al. (2017) and Sihag et al. (2019a) showed successful utilization of ANFIS in modelling the cumulative infiltration and the unsaturated hydraulic conductivity of soil samples. Some latest studies suggested successful application of soft computing techniques, viz. SVM, GPR, M5 tree and random forest regression to the field of groundwater hydrology (Singh et al. 2017, 2019a, b; Angelaki et al. 2018; Sihag et al. 2018a, b, c; Vand et al. 2018; Kumar and Sihag 2019; Sihag et al. 2019b, c), water resources (Kumar et al. 2018; Sepahvand et al. 2019; Singh et al. 2018a, b; Tiwari and Sihag 2018; Tiwari et al. 2019) and engineering (Nain et al. 2018, 2019; Mehdipour et al. 2018; Kumar et al. 2019; Mohanty et al.

2019). Keeping in view the importance of M5 tree and random forest regression techniques, the present research deals with the implementation of these techniques in an attempt to relate unsaturated hydraulic conductivity of the field data measured from 20 locations of Kurukshetra district, Haryana, with the soil physical properties.

To the best knowledge of authors, the predictive capabilities of M5 tree and random forest (RF) regression are not investigated in estimating the unsaturated hydraulic conductivity of soil in the field. So this study investigates the potential of M5 tree and RF regression models. A relationship based on multiple nonlinear regression (MNLR) is developed for the unsaturated hydraulic conductivity of soil considering sand (%), clay (%), silt (%), bulk density and moisture content as input variables, and the developed relationship is compared with the soft computing-based regression models (M5 and RF).

Study area

Kurukshetra district lies in the Ghaggar basin (Fig. 1), and it is in the north-east part of the Haryana State, India. Thanesar Tehsil of Kurukshetra district is chosen for experimentation. Ghaggar is one of the main rivers of Haryana State, India. Twenty different locations were selected for measurement of infiltration process. The texture of the soil is listed and shown in Table 1 and Fig. 2, respectively.

Data set

The unsaturated soil hydraulic conductivity was measured in the field using a mini disc infiltrometer (Decagon Devices Inc.) as shown in Fig. 3. During the experiment, the volume of water in the lower chamber was listed at expected time intervals. The total data set consisting 240 observations from field experiments of infiltration process was separated randomly into two groups of training and testing, respectively. Larger group is considered as training data (70% of the total data), while smaller group is considered as testing data (rest 30% of the total data). Input parameters are sand, clay, silt, bulk density and moisture content, and output parameter is unsaturated hydraulic conductivity (K) of soil. The characteristics of both data sets are listed in Table 2.

Modelling approaches

Multiple nonlinear regression (MNLR)

To develop nonlinear regression model, the general form of multiple nonlinear regression model is considered by the following relationship:
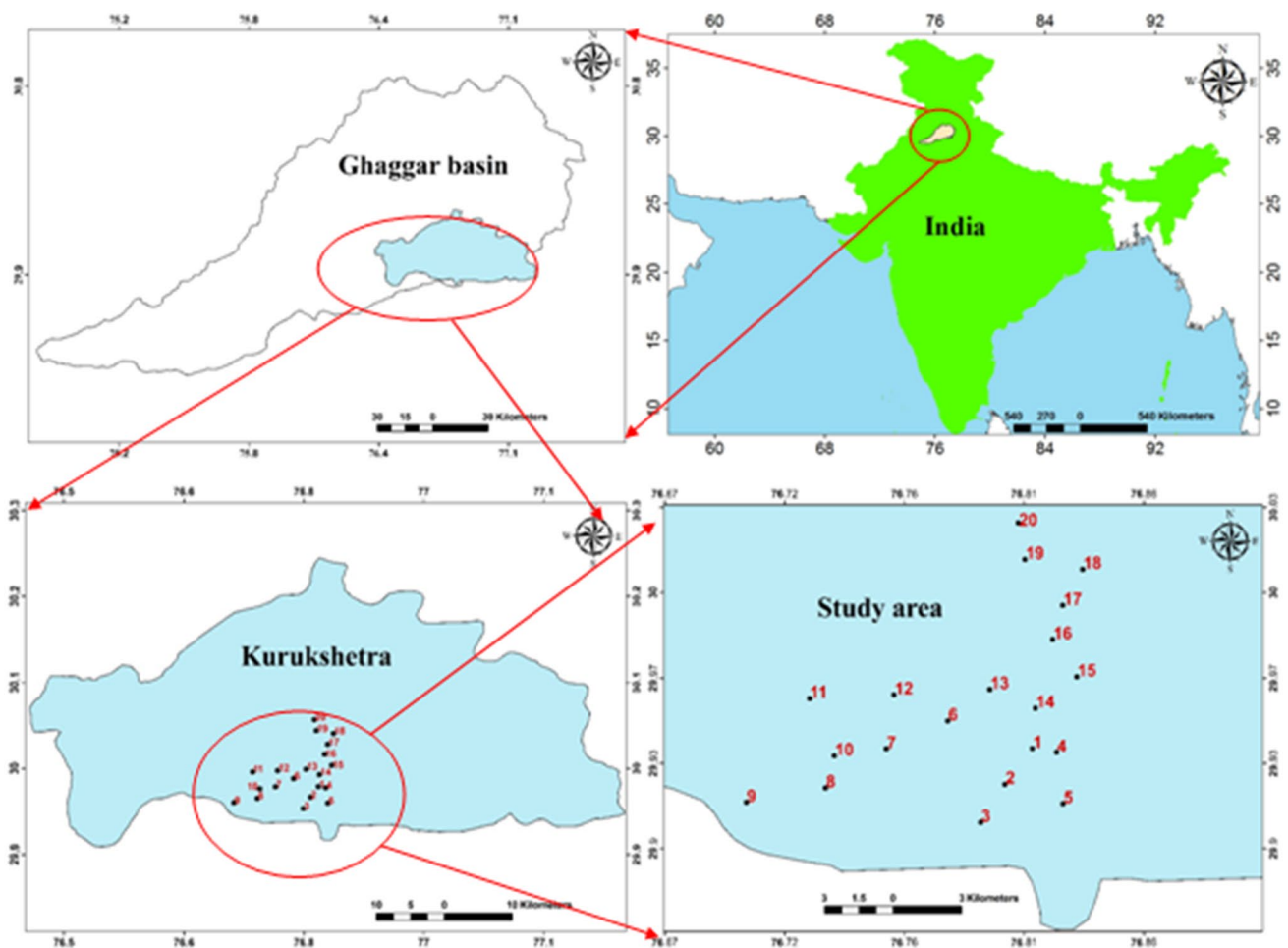
Fig. 1 Study area

$$K = aS^{b_1}C^{b_2}\text{Si}^{b_3}\rho^{b_4}\text{MC}^{b_5} \qquad (1)$$

where K is the dependent variable representing hydraulic conductivity of soil, S, C, Si, ρ and MC are regarded as explanatory variables, a is the constant, and the estimate of parameters (regression coefficients) b_1, b_2, b_3, b_4, and b_5 is obtained by minimizing the sum of squares of error in prediction based on least squares. Based on the above equation, the following relationship is developed from the training data set:

$$K = 0.83 \times 10^{-5}S^{0.95}C^{-0.23}\text{Si}^{-0.37}\rho^{1.92}\text{MC}^{-0.015} \qquad (2)$$

M5 model tree (M5)

M5 tree, introduced by Quinlan (1992), is a decision tree learner for regression problems. This tree algorithm assigns linear regression functions at the terminal nodes and fits a multivariate linear regression model to each subspace by classifying or dividing the whole data space into several sub spaces. The M5 tree method deals with continuous class problems instead of discrete classes and can handle tasks with very high dimensionality. It reveals piecewise information of each linear model constructed to approximate nonlinear relationships of the data set.

The information about the splitting criteria for the M5 model tree is gained on the basis of calculates of error at each node. The error is analysed by the standard deviation of the class values that arrive at a node. The attribute that maximizes the expected error reduction resulting from the testing of each attribute at that node is chosen for splitting at the node. The standard deviation reduction (SDR) is calculated by:

$$\text{SDR} = \text{sd}(K) - \sum \frac{|K_i|}{|K|}\text{sd}(K_i) \qquad (3)$$

where K indicates set of instances that attain the node; K_i indicates the subset of illustrations that have the ith product of the possible set; and sd indicates the standard deviation.

Table 1 Texture of the soil

Site no.	Location	Texture	Sand (%)	Clay (%)	Silt (%)
1	Dayalpur	Loamy sand	78.73	7.4445	13.8255
2	Samshipur	Clay	39.84	55.3472	4.8128
3	Kirmach (SKS)	Clay	37.14	43.3734	19.4866
4	Alampur	Sandy clay loam	47.5	25.2	27.3
5	Sanheri Khalsa	Sandy clay loam	52.11	24.9028	22.9872
6	Mirzapur	Clay	26.63	41.8209	31.5491
7	Khanpur Roran	Clay loam	32.94	29.5064	37.5536
8	Barna	Clay loam	31.52	35.7133	32.7667
9	Pindarsi	Sandy clay loam	47.6	27.248	25.152
10	Kamoda	Loam	42.85	24.003	33.147
11	Lohar Majra	Clay loam	24.6	39.962	35.438
12	Jyotisar	Sandy clay loam	52.71	34.5217	12.7683
13	Narkatari	Clay loam	22.93	32.3694	44.7006
14	Kurukshetra University	Clay	52.74	19.85	27.41
15	Thim Park	Clay	36.7	26.586	36.714
16	Darra Khera	Sandy clay loam	35.31	59.5148	5.1752
17	Bhiwani Khera	Sandy clay loam	59.58	30.7192	9.7008
18	Bahadur Pura	Clay	50.78	23.6256	25.5994
19	Hansala	Loam	19.74	62.6028	17.6572
20	Durala	Sandy loam	39.13	46.2612	14.6088

Fig. 2 Texture of the soil for the study area
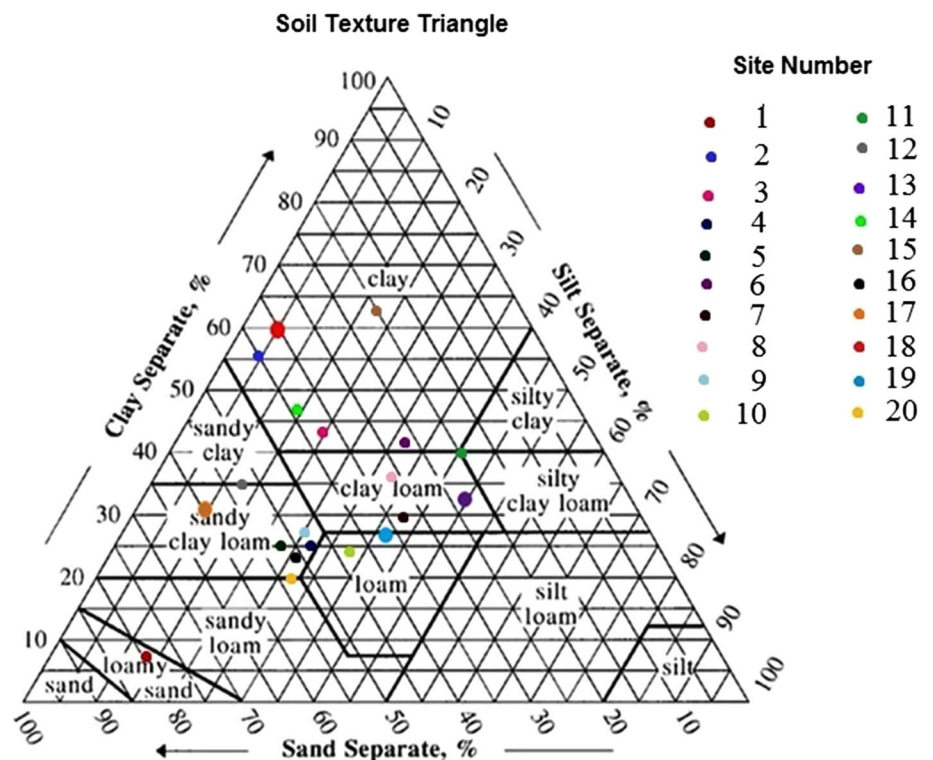
Random forest regression (RF)

RF regression approach was initially introduced by Breiman (2001). This is a machine learning classifier that contains several decision trees and targets the class that is the mode of the classes' target by individual trees. Number of trees to be grown (k) in the forest and the quantity of features or variables chosen (m) at every node to develop a tree are the
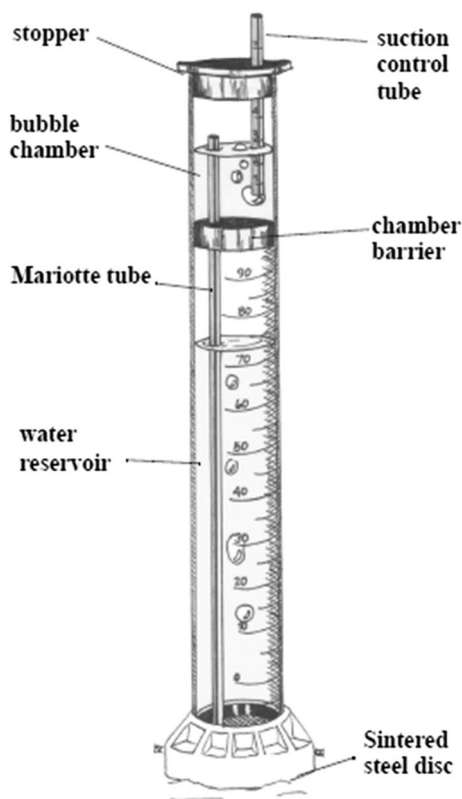
Fig. 3 Mini disc infiltrometer (Infiltrometer User's Manual, 2014)

two standard user-defined parameters required for random forest regression (Breiman 2001). In this study, we applied RF model to predict the unsaturated hydraulic conductivity of soil (K).

Implementation of machine learning methods

Three standard statistical measures: coefficient of determination (R^2), root mean square error (RMSE) and mean

absolute error (MAE), were implemented as performance evaluation parameters in order to judge the performance of the machine learning methods. Large numbers of models were developed by changing the user-defined parameters of the modelling approaches with the training data set, and the efficiency of the developed models was validated by analysing the performance on the testing data set. So the modelling procedure involves three steps: changing the values of user-defined parameters associated with the modelling method, model building on training data and validation of the formed models on testing data. The identification of model-specific user-defined parameters and the selection of the models were based on statistical criteria. Higher values of R^2 and lower values of RMSE and MAE indicate better estimation accuracy of the models. Number of trees to be grown (k) in the forest and the number of features or variables selected (m) at each node to generate a tree are the two standard user-defined parameters required for random forest regression. In M5 tree model, calibration of models was done by means of changing the value of no. of instances allowed at each node. The selected primary parameters of the modelling approaches are presented in Table 3.

Results and discussion

The efficiency of the modelling methods in predicting the hydraulic conductivity of soil in the field is tested by developing the models by regression modelling methods and testing the accuracy of the developed models with the unseen

Table 3 Primary parameters

Machine learning approach	Primary parameters
M5 model tree (pruned and unpruned)	Instances = 4
Random forest	$k = 100$, $m = 10$

Table 2 Characteristics of training and testing data sets

Parameter	Unit	Data set	Lower	Higher	Mean	SD
Sand (S)	(%)	Training	19.74	78.730	41.89	13.95
		Testing	19.74	78.73	40.75	13.69
Clay (C)	(%)	Training	7.44	62.603	34.17	13.44
		Testing	7.44	62.60	35.40	13.91
Silt (Si)	(%)	Training	4.81	44.70	23.95	10.89
		Testing	4.81	44.70	23.86	11.67
Bulk density (ρ)	g/cc	Training	1.39	1.89	1.68	0.13
		Testing	1.39	1.89	1.66	0.13
Moisture content (MC)	(%)	Training	0.37	16.87	4.86	3.69
		Testing	0.28	17.12	4.53	3.60
Unsaturated hydraulic conductivity (K) $\times 10^{-5}$	cm/s	Training	0.055	60.87	12.69	11.9
		Testing	0.04	57.14	12.73	11.57

testing data. The inputs selected for estimating the hydraulic conductivity are sand (%), clay (%), silt (%), bulk density and moisture content. The performance of multiple nonlinear regression (MNLR) is evaluated by generating a simple multivariate relationship (Eq. 2) based on nonlinear regression function (Eq. 1) applied to the training data set. In order to check the potential of the nonlinear relationship (Eq. 2), the equation is applied to the testing data set and the outcomes are depicted in Fig. 4 as a scattering diagram of the predicted data of hydraulic conductivity. Closeness of the data to the perfect agreement line represents accuracy of the model in estimating the actual field data. However, in this case, excessive scattering of the data points from the agreement line reveals poor performance of the MNLR model in approximating the actual data of field hydraulic conductivity and hence lacking in generalization. The statistical measures observed with the testing data verify the lower accuracy of the MNLR modelling technique as the error values (RMSE and MAE) are higher and the coefficient of determination (R^2) is less (Table 4). So a direct relationship is not sufficient to precisely relate the hydraulic conductivity with the soil input parameters used in the current study, leading to inferior performance by the MNLR model.

In an attempt to approximate the actual field data of hydraulic conductivity of soil, machine learning methods are adopted to improve the generalization capacity. M5 model tree algorithm, which utilizes linear regression models to define input–output relationship based on splitting of the parameter space of the data set into several subspaces, was used. Two M5 tree models: pruned and unpruned trees, were developed by changing the instances used at the leaf node. The values of user-defined parameters (instances used) were selected by implementing M5 model tree method on the training data and judging the performance on the testing data (Table 3). By checking the results of both pruned and unpruned tree models with the testing data set, the statistical measures indicate lower values of RMSE (0.0000699) and MAE (0.0000488) obtained with unpruned M5 tree model relative to pruned (RMSE = 0.0000898, MAE = 0.0000633) tree model. The higher value of R^2 observed with unpruned model infers closer prediction of actual data, and scattering plot shows (Fig. 5) that the estimated points of the unpruned model lie closer to the agreement line when compared with the pruned model tree. So based on the results, unpruned model indicates better learning capability than pruned model as the estimation accuracy is higher.

The development of random forest model is achieved by carrying out trials with the training data set by changing the number of features used at each node to generate a tree, and the numbers of trees and finally the performance of the calibrated model are tested on the testing data set. After optimizing the performance of the testing data by checking the forecasting accuracy of the developed model based on least RMSE and MAE values, the model was selected based on generalization ability. The performance of RF regression is presented in Fig. 6 as a comparison of actual and predicted
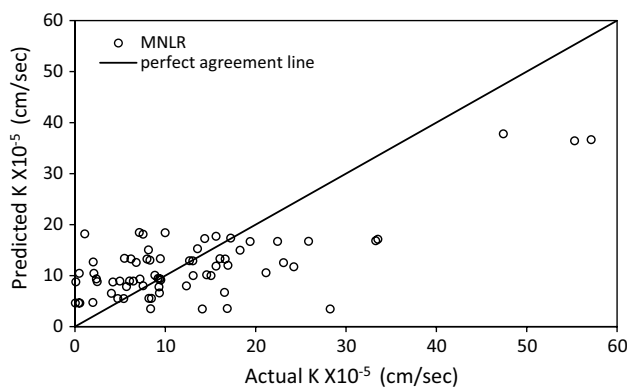

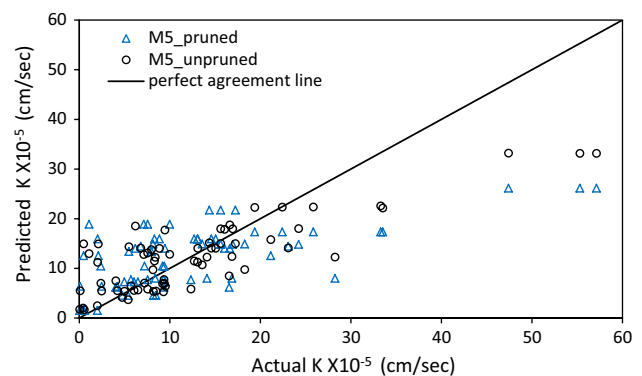**Fig. 4** Actual versus predicted values of hydraulic conductivity using MNLR model (testing data)

Fig. 5 Actual versus predicted values of hydraulic conductivity using M5 model tree (testing data)

Table 4 Statistical performance measures for modelling approaches

Approach	Training data			Testing data		
	R^2	RMSE	MAE	R^2	RMSE	MAE
MNLR (Eq. 2)	0.294	0.0001018	0.0000783	0.530	0.0000807	0.0000611
M5_pruned	0.253	0.0001022	0.0000792	0.402	0.0000898	0.0000633
M5_unpruned	0.515	0.0000836	0.0000619	0.675	0.0000699	0.0000488
RF	0.90	0.0000409	0.0000297	0.819	0.0000491	0.0000396
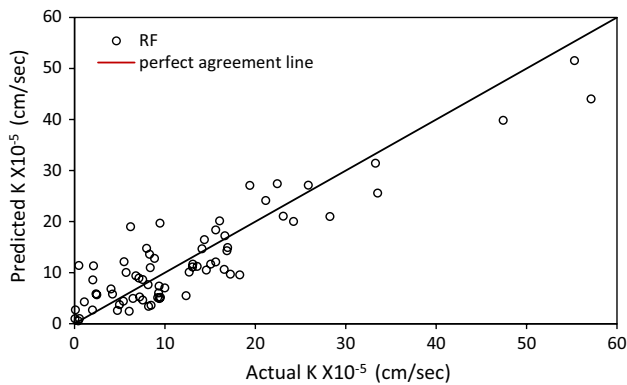
Fig. 6 Actual versus predicted values of hydraulic conductivity using RF model (testing data)

Fig. 7 Comparison of soft computing models in estimating the testing data

values of hydraulic conductivity. It is analysed from the plot that the scattering of the data is relatively closer to the perfect agreement line. The RF model generated comparatively lower values of RMSE (0.0000491) and MAE (0.0000396) than the other tested regression models (Table 4), which indicates the superior potential of the RF model in accurately relating the hydraulic conductivity of the field data with the soil properties.

Comparative analysis of the regression models

The efficacy of MNLR, M5 tree and RF regression in estimating the hydraulic conductivity of field data is tested and presented as a combined graph showing all the applied regression models (Fig. 2). To study the scatter around the perfect agreement line, the graph between actual and predicted values is represented by error lines in the range of $\pm 30\%$. From Fig. 7, it is clear that the prediction performance of the random forest (RF) model is well within error range of $\pm 30\%$ except for some smaller values. The model measures the actual data with an accuracy of $\pm 30\%$. Lower values of RMSE and MAE obtained with RF model confirm this (Table 4). The scattering of the MNLR model from the perfect agreement line is higher (except for some larger values) than all the other models indicating inferior performance of the model in estimation and generalization. Both M5 tree models overpredict the smaller values of hydraulic conductivity and reside outside the $+30\%$ error line, but underpredict for the larger values and lie near to the -30% error line. The scattering of the M5_unpruned model is relatively more than that of M5_pruned model indicating better performance by the unpruned M5 tree model. So based on statistical measures and error plots, the performance of RF model is found superior to M5 model tree and nonlinear regression model.

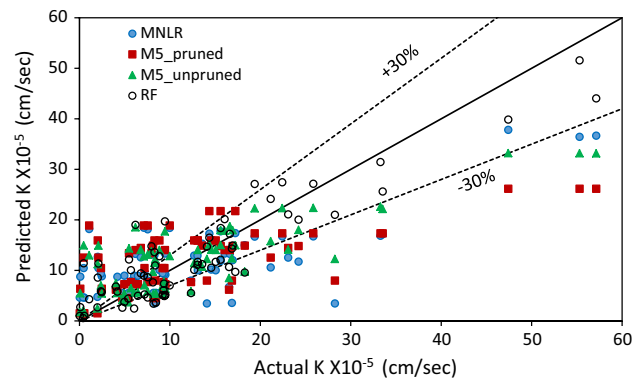To analyse the relative variation of the implemented modelling techniques and the actual experimental field data, a graph between the number of observations and hydraulic conductivity of the field is presented (Fig. 8). This figure shows that RF based model follows the same path as followed by actual observed hydraulic conductivity values so RF model is most suitable for estimating the hydraulic conductivity of soil than other above discussed models. The deviation of the predicted points from the actual points by M5_pruned model is the highest from all the models.

As shown in Fig. 9, the RF model significantly reduces the overall residual errors due to accurate predictions by the model. Other regression models have larger residuals than RF model, thus indicating low efficiency of the models in accurate estimations of the field data.

ANOVA test using single factor was used to compare the statistical significance of predicted values from machine learning approaches and actual values. Results suggest that F-value was less than the F-critical and P value was greater than 0.05 for all the soft computing models which indicate that the difference in predicted and actual values was insignificant (Table 5).

Conclusions

Machine learning methods are employed for the purpose of accurate and reliable predictions of hydraulic conductivity of soil. Twenty different locations in the district of Kurukshetra, Haryana (India), were selected for the experimental data collection on monthly basis for the period of 1 year. Mini disc infiltrometer was used for the determination of hydraulic conductivity in the field. The compiled field data of hydraulic conductivity associated with soil physical properties: sand (%), clay (%), silt (%), bulk density and moisture content as input parameters, were used for modelling by the random division of the total data in two parts (training and testing). The modelling techniques employed in this study were multivariate nonlinear regression, M5

Fig. 8 Scattering of actual and soft computing model-based predictions against observation number
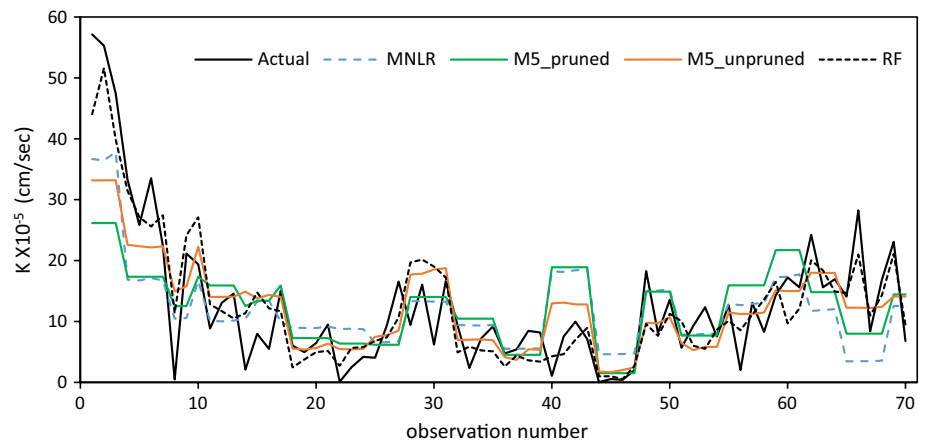

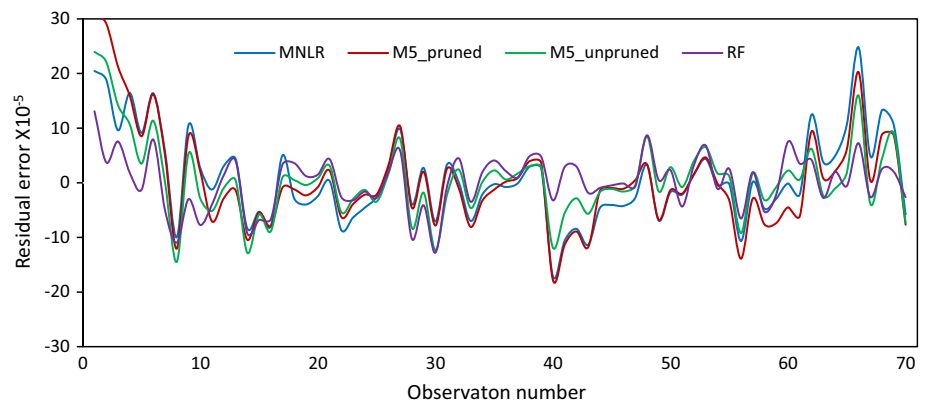**Fig. 9** Residual error versus observation number using soft computing models

Table 5 Result of ANOVA single-factor test

Model	F-value	P value	F-critical	Difference
MNLR (Eq. 2)	0.24088	0.62435	3.909729	Insignificant
M5_pruned	0.047406	0.827962	3.909729	Insignificant
M5_unpruned	0.081122	0.776209	3.909729	Insignificant
RF	0.012373	0.911592	3.909729	Insignificant

model tree and random forest (RF) regression. Based on the validation results of the developed regression models on the testing data set, the performance of RF regression in predicting the hydraulic conductivity of field data was found more accurate than M5 model tree as well as the relationship developed on the basis of multiple nonlinear regression. The performance of unpruned M5 tree model is found superior to both pruned M5 tree and multiple nonlinear regression models. The modelling results based on standard statistical measures indicated that the RF model, due to higher predictive efficiency in model development and validation, has higher generalization capability and thus can be applied for the accurate estimations of the field hydraulic conductivity of soil relating to basic soil properties.

Compliance with ethical standards

Conflict of interest Parveen Sihag, Sahar Mohsenzadeh Karimi and Dr. Anastasia Angelaki declared that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Agyare WA, Park SJ, Vlek PLG (2007) Artificial neural network estimation of saturated hydraulic conductivity. Vadose Zone J 6(2):423–431

Al-Sulaiman M, Aboukarima A (2016) Prediction of unsaturated hydraulic conductivity of agricultural soils using artificial neural network and c#. J Agric Ecol Res Int 5(4):1–15. https://doi.org/10.9734/jaeri/2016/21622

Angelaki A, Singh Nain S, Singh V, Sihag P (2018) Estimation of models for cumulative infiltration of soil using machine learning

methods. ISH J Hydraul Eng. https://doi.org/10.1080/09715010.2018.1531274.

Arshad RR, Sayyad G, Mosaddeghi M, Gharabaghi B (2013) Predicting saturated hydraulic conductivity by artificial intelligence and regression models. ISRN Soil Sci. https://doi.org/10.1155/2013/308159

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Das SK, Samui P, Sabat AK (2012) Prediction of field hydraulic conductivity of clay liners using an artificial neural network and support vector machine. Int J Geomech 12(5):606–611. https://doi.org/10.1061/(asce)gm.1943-5622.0000129

Ekhmaj AI (2010) Predicting soil infiltration rate using artificial neural network. In: 2010 International conference on environmental engineering and applications (ICEEA), pp 117–121. IEEE

Elbisy MS (2015) Support Vector Machine and regression analysis to predict the field hydraulic conductivity of sandy soil. KSCE J Civil Eng 19(7):2307–2316

Erzin Y, Gumaste SD, Gupta AK, Singh DN (2009) Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soils. Can Geotech J 46(8):955–968. https://doi.org/10.1139/t09-035

Jarvis N, Koestel J, Messing I, Moeys J, Lindahl A (2013) Influence of soil, land use and climatic factors on the hydraulic conductivity of soil. Hydrol Earth Syst Sci 17(12):5185–5195

Kumar M, Sihag P (2019) Assessment of Infiltration rate of soil using empirical and machine learning-based models. Irrigation and Drainage, Wiley. https://doi.org/10.1002/ird.2332

Kumar M, Tiwari NK, Ranjan S (2018) Prediction of oxygen mass transfer of plunging hollow jets using regression models. ISH J Hydraul Eng. https://doi.org/10.1080/09715010.2018.1435311

Kumar M, Sihag P, Singh V (2019) Enhanced soft computing for ensemble approach to estimate the compressive strength of high strength concrete. J Mater Eng Struct 6(1):93–103

Mehdipour V, Stevenson DS, Memarianfard M, Sihag P (2018) Comparing different methods for statistical modeling of particulate matter in Tehran, Iran. Air Qual Atmos Health 11(10):1155–1165

Mohanty S, Roy N, Singh SP, Sihag P (2019) Estimating the strength of stabilized dispersive soil with cement clinker and fly ash. Geotech Geol Eng. https://doi.org/10.1007/s10706-019-00808-1

Nain SS, Sihag P, Luthra S (2018) Performance evaluation of fuzzy-logic and BP-ANN methods for WEDM of aeronautics super alloy. MethodsX 5:890–908

Nain SS, Garg D, Kumar S (2019) Modelling and analysis for the machinability evaluation of Udimet-L605 in wire-cut electric discharge machining. Int J Process Manag Benchmark 9(1):47–72

Quinlan JR (1992) Learning with continuous classes. In: Adams S (ed) Proceedings of AI'92. World Scientific, Singapore, pp 343–348

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A (2012) Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. Math Geosci 44(6):739–763. https://doi.org/10.1007/s11004-012-9409-2

Sepahvand A, Singh B, Sihag P, Nazari Samani A, Ahmadi H, Fiz Nia S (2019) Assessment of the various soft computing techniques to predict sodium absorption ratio (SAR). ISH J Hydraul Eng. https://doi.org/10.1080/09715010.2019.1595185

Sihag P (2018) Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. Model Earth Syst Environ 4(1):189–198

Sihag P, Tiwari NK, Ranjan S (2017a) Estimation and inter-comparison of infiltration models. Water Sci 31(1):34–43

Sihag P, Tiwari NK, Ranjan S (2017b) Modelling of infiltration of sandy soil using gaussian process regression. Model Earth Syst Environ 3(3):1091–1100

Sihag P, Jain P, Kumar M (2018a) Modelling of impact of water quality on recharging rate of storm water filter system using various kernel function based regression. Model Earth Syst Environ. https://doi.org/10.1007/s40808-017-0410-0

Sihag P, Singh B, Gautam S, Debnath S (2018b) Evaluation of the impact of fly ash on infiltration characteristics using different soft computing techniques. Appl Water Sci 8(6):187

Sihag P, Tiwari NK, Ranjan S (2018b) Prediction of cumulative infiltration of sandy soil using random forest approach. J Appl Water Eng Res 7(2):118–142. https://doi.org/10.1080/23249676.2018.1497557

Sihag P, Tiwari NK, Ranjan S (2019a) Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). ISH J Hydraul Eng 25(2):132–142

Sihag P, Esmaeilbeiki F, Singh B, Pandhiani SM (2019b) Model-based soil temperature estimation using climatic parameters: the case of Azerbaijan Province, Iran. Geol Ecol Landscapes. https://doi.org/10.1080/24749508.2019.1610841

Sihag P, Esmaeilbeiki F, Singh B, Ebtehaj I, Bonakdari H (2019c) Modeling unsaturated hydraulic conductivity by hybrid soft computing techniques. Soft Comput. https://doi.org/10.1007/s00500-019-03847-1

Singh B, Sihag P, Singh K (2017) Modelling of impact of water quality on infiltration rate of soil by random forest regression. Model Earth Syst Environ 3(3):999–1004

Singh B, Sihag P, Singh K (2018a) Comparison of infiltration models in NIT Kurukshetra campus. Appl Water Sci 8(2):63. https://doi.org/10.1007/s13201-018-0708-8

Singh B, Sihag P, Singh K, Kumar S (2018b) Estimation of trapping efficiency of a vortex tube silt ejector. Int J River Basin Manag. https://doi.org/10.1080/15715124.2018.1476367

Singh B, Sihag P, Deswal S (2019a) Modelling of the impact of water quality on the infiltration rate of the soil. Appl Water Sci 9(1):15. https://doi.org/10.1007/s13201-019-0892-1

Singh B, Sihag P, Pandhiani SM, Debnath S, Gautam S (2019b) Estimation of permeability of soil using easy measured soil parameters: assessing the artificial intelligence-based models. ISH J Hydraul Eng. https://doi.org/10.1080/09715010.2019.1574615

Tiwari NK, Sihag P (2018) Prediction of oxygen transfer at modified Parshall flumes using regression models. ISH J Hydraul Eng. https://doi.org/10.1080/09715010.2018.1473058

Tiwari NK, Sihag P, Ranjan S (2017) Modeling of infiltration of soil using adaptive neuro-fuzzy inference system (ANFIS). J Eng Technol Educ 11(1):13–21

Tiwari NK, Sihag P, Singh BK, Ranjan S, Singh KK (2019) Estimation of tunnel desilter sediment removal efficiency by ANFIS. Iran J Sci Tech Trans Civ Eng. https://doi.org/10.1007/s40996-019-00261-3

Vand AS, Sihag P, Singh B, Zand M (2018) Comparative evaluation of infiltration models. KSCE J Civil Eng 22(10):4173–4184

Springer